Georgia State University
ScholarWorks @ Georgia State University

AYSPS Dissertations | Andrew Young School of Policy Studies

5-14-2021

ESSAYS ON ENVIRONMENTAL AND HEALTH ECONOMICS

Golnoush Rahimzadeh
grahimzadeh1@student.gsu.edu

Follow this and additional works at: https://scholarworks.gsu.edu/aysps_dissertations

Recommended Citation

Rahimzadeh, Golnoush, "ESSAYS ON ENVIRONMENTAL AND HEALTH ECONOMICS." Dissertation, Georgia State University, 2021.
https://scholarworks.gsu.edu/aysps_dissertations/26

This Dissertation is brought to you for free and open access by the Andrew Young School of Policy Studies at ScholarWorks @ Georgia State University. It has been accepted for inclusion in AYSPS Dissertations by an authorized administrator of ScholarWorks @ Georgia State University. For more information, please contact scholarworks@gsu.edu.

ABSTRACT

ESSAYS ON ENVIRONMENTAL AND HEALTH ECONOMICS

By

GOLNOUSH RAHIMZADEH

MAY 2021

Committee Chair: Dr. Henry Spencer Banzhaf

Major Department: Economics

In this dissertation, I study two questions in environmental and health economics with a focus on the role of urban configuration. The first essay investigates the impact of urban sprawl on the temperature in the United States. Sprawl contributes to the heat island effect by eliminating vegetation, expanding dark surfaces, and increasing daily travel distance. This study quantifies this effect by constructing and linking the required measures and exploiting variations in the data using different identification strategies. I construct an index of residential compactness in US metropolitan areas using satellite remote sensing information to analyze landscape changes from 1974–2012 and link them to the Global Surface Summary of the Day data. To address the reverse causality issue, I utilize the planned interstate highways emanating from the central cities as an instrument for sub-urbanization in the United States. I also examine the Impact of Sprawl on UHI by introducing a control group for each metropolitan statistical area (MSA) in the sample. The results suggest a positive and causal relationship between the temperature of the MSA center and urban sprawl. Thus, horizontal development of the city imposes an extra burden on the temperature of the city center.

The second essay, which is joint work with Dr. Firouzi Naeim, studies the role of labor unions in response to the pandemic. Labor unions are among the largest institutions in the United States, and their role in regulating employee-employer relations is hard to ignore. Costly efforts to control the spread of COVID-19, combined with the monopoly and collective voice faces of unions, emphasize the role unions can play in shaping the response of the

workforce in coping with COVID-19. We analyze the effect of union size by utilizing state-level data in the United States and by employing a dynamic nonlinear probability model. The results suggest new evidence of positive externalities for union employees compared with nonunion employees. We find that increasing union size by 1,000 new members in the United States would lead to 110 fewer COVID-19 cases 11 months after the onset of the virus, controlling for hours of work and differences in union members' characteristics.

ESSAYS ON ENVIRONMENTAL AND HEALTH ECONOMICS

BY

GOLNOUSH RAHIMZADEH

A Dissertation Submitted in Partial Fulfillment
of the Requirements for the Degree
of
Doctor of Philosophy
in the
Andrew Young School of Policy Studies
of
Georgia State University

GEORGIA STATE UNIVERSITY
2021

Copyright by
Golnoush Rahimzadeh
2021

ACCEPTANCE

This dissertation was prepared under the direction of the candidate's Dissertation Committee. It has been approved and accepted by all members of that committee, and it has been accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Economics in the Andrew Young School of Policy Studies of Georgia State University.

Dissertation Chair: Dr. H. Spencer Banzhaf

Committee: Dr. Garth Heutel
Dr. Carlianne Patrick
Dr. Maryam Naghsh Nejad

Electronic Version Approved:

Dr. Sally Wallace, Dean
Andrew Young School of Policy Studies
Georgia State University
May, 2021

DEDICATION

To My Mother and Father.

ACKNOWLEDGEMENTS

I am beyond thankful to my advisor, Spencer Banzhaf, for his continuous support and guidance throughout the completion of this research. I am immensely privileged to have him as my academic mentor. I'm grateful to my committee members, Garth Heutel, Carlianne Patrick, and Maryam Naghsh Nejad, for their support and constructive input throughout this process. I thank Juha Siikamaki, Daniel Kreisman, Tom Mroz, and Barry Hirsch for their insightful feedback. I would also like to thank the economics faculty and staff at Georgia State for their effort and commitment. I would like to extend my gratitude to Rusty Tchernis for always being supportive and encouraging. I'm also thankful for the participants' comments in the GSU Ph.D. seminar series, Southern Economics Association, and Econometrics Society. Finally, I give a special thank you to Peyman Firouzi Naeim, my friend, husband, and professional fellow, for his unwavering support.

Contents

Chapter I The Impact of Urban Sprawl on Temperature in the United States **1**

1.1 Introduction 1

1.2 Measuring Urban Sprawl 4

1.2.1 Residential Compactness 7

1.2.2 MSA Center 9

1.3 Estimation 10

1.3.1 Identification Strategy 10

1.4 Data 17

1.4.1 Temperature Data 17

1.4.2 Assigning Temperature to MSA Center 17

1.4.3 Setting Bounds 20

1.5 Estimation Results 21

1.5.1 Fixed and Flexible Bounds 21

1.5.2 OLS Estimates 24

1.5.3 Correction for Unobserved Fixed Heterogeneity 28

1.5.4 Trend Analysis Results 29

1.5.5 The Impact of Sprawl on Surrounding Urban Area Using Instrumental Variable 30

1.5.6 The Impact of Sprawl on UHI 36

1.6 Robustness Check: Subsample Estimation 38

1.7 Conclusion 40

Chapter II The Role of Labor Unions in Response to COVID-19 Pandemic **42**

2.1 Introduction 42

2.2 Mechanisms 44

2.3 Model and Estimation 48

2.3.1 Introduction to Model 48
2.3.2 Identification 53
2.4 Data 55
2.5 Results 58
2.5.1 Employment Dynamics 66
2.5.2 Discussion 68
2.6 Robustness Check: OLS and Median Regression Results 69
2.7 Conclusion 70

Appendix A. Appendix for Chapter I **73**
A.1 Additional tables 73
A.2 Additional Figures 83

Appendix B. Appendix for Chapter II **86**
B.1 Additional tables 86

Bibliography **90**

Vita **95**

List of Tables

Table 1.1 Availability of the weather stations in GSOD data 18

Table 1.2 Calibrated values to being used in flexible bound scheme 21

Table 1.3 Number of MSA's and average distance between center and In- and out-bound stations 24

Table 1.4 OLS results of the effects on mean Temperature 28

Table 1.5 FE OLS results of the effects on mean Temperature 29

Table 1.6 Diff-in-Diff effects on mean Temperature 31

Table 1.7 F-stat of First stage of IV 33

Table 1.8 IV effects on mean Temperature of urban area 34

Table 1.9 IV effect on mean Temperature of rural area (Control group) 35

Table 1.10 IV effects on UHI 37

Table 1.11 IV result of Effects on mean Annual Temperature for subsample of data 40

Table 2.12 Summary of statistics. Socio-economic ($\vec{X}$) and teartment ($\vec{T}$) variables 57

Table 2.13 Summary of statistics. Fixed in time variables ($\vec{\Gamma}$) 58

Table 2.14 Simulated marginal effects 63

Table 2.15 Effect of controlling for the simultaneity and selection bias. 65

Table 2.16 Sensitivity of OLS estimates to the odd ratios close to zero. 70

Table 2.17 Sensitivity of median regression to the odd ratios close to zero. 71

Table A.1 OLS results of the effects for multiple outcomes (Fixed Scheme) 73

Table A.2 Difference-in-Difference effects for multiple outcomes (Fixed Scheme) . 74

Table A.3 IV effects for multiple outcomes in urban area (Fixed Scheme) 75

Table A.4 IV effects for multiple outcomes in surrounding area (Fixed Scheme) . 76

Table A.5 Causal effect of residential compactness on difference between urban and surrounding area for multiple outcomes (Fixed Scheme) 77

Table A.6 OLS results of the effects for multiple outcomes (Flexible Scheme) . . 78

Table A.7 Difference-in-Difference effects for multiple outcomes(Flexible Scheme) 79
Table A.8 Instrumental Variable estimates for multiple outcomes (Flexible Scheme) 80
Table A.9 Analysis of the sensitivity of the rural area to the Instrument for multiple outcomes(Flexible Scheme) 81
Table A.10 Causal effect of residential compactness on difference between urban and surrounding area for multiple outcomes 82

Table B.1 Estimated MM coefficients of a logistic model. 86
Table B.2 Estimated MM coefficients of a logistic model. 87
Table B.3 Estimated MM coefficients of a logistic model. 88
Table B.4 Applied restrictions in states and DC. 89

List of Figures

Figure 1.1 Population growth and geographic distribution in the United States . 2

Figure 1.2 Construction of Residential Compactness and Commercial Accessibility ratios for Atlanta metro area 11

Figure 1.3 Finding Central Business District (CBD) and setting the city bound . 12

Figure 1.4 Average distance from MSA center to inner city and outer city weather station 23

Figure 1.5 MSA size and Residential Compactness 25

Figure 1.6 Movements of the temperature in time (Fixed bounds) 26

Figure 1.7 Movements of the temperature in time (Flexible bounds) 27

Figure 1.8 Distributional effect of Treatment and Time in DiD 32

Figure 1.9 MSA size and Residential Compactness 39

Figure 2.10 Number of new COVID-19 cases (observed and simulated) 56

Figure 2.11 Total COVID-19 cases (observed and simulated) 60

Figure 2.12 Dynamic change in total number of COVID-19 cases due to change in each treatment. 64

Figure 2.13 Employment dynamics in 2020 67

Figure A.1 GSOD data of Annual Temperature 83

Figure A.2 GSOD data of Visibility 84

Figure A.3 GSOD data of Dew Point 85

Chapter I

The Impact of Urban Sprawl on Temperature in the United States

1.1 Introduction

Growing population and increasing temperature are two trends that have had a profound impact on communities. In addition, geographical distribution of the population is changing in favor of urbanization. In the United States, the urban population almost doubled from 126 million in 1960 to 249 million in 2010. During the same period, the rural population grew from 54 million to 59 million, barely a 10 % increase (Figure 1.1). The need for new urban areas means policymakers must choose between horizontal versus vertical development of cities. One of the areas of concern is the effect of each of the alternatives on the environment and, in particular, the temperature of cities; metropolitan areas are significantly warmer than their surrounding rural areas, which produces a phenomenon known as the urban heat island (UHI). The UHI affects "summertime peak energy demand, air conditioning costs, air pollution and greenhouse gas emissions, heat-related illness and mortality, and water pollution".[1]. Over the past five decades, the average temperature in the United States has increased and is expected to continue to rise.[2]

While a rise in temperature can be comforting in cities with a cold climate or in cold seasons, in a relatively warmer climate and in warmer months of the year, a rise in temperature

[1] United States Environmental Protection Agency

[2] The annual average temperature of the contiguous United States has risen since the start of the 20th century. In general, the temperature increased until about 1940, decreased until about 1970, and increased rapidly through to 2016. In particular, annual average temperature over the contiguous United States has increased by $1.2°F(0.7°C)$ for the period 1986–2016 relative to 1901–1960. There is general consistency between surface and satellite data in their depiction of the rapid warming of the past few decades. The annual average temperature of the contiguous United States is projected to rise throughout the present century. Increases for the period 2021–2050 relative to 1976–2005 are projected to be between $2.5°F$ and $2.9°F$ (Zhongming et al. (2017))

has an undesirable and harmful effect on human health, economic productivity, and energy consumption. Urban climate features, such as temperature, are affected by urban structure. A positive relationship between soil sealing and land surface temperature has been detected in many studies (Weng et al. (2007), Schueler (1994)). This relationship suggests that urban sprawl can elevate the UHI effect both in geographic extent and intensity (Bhatta (2010)). Extending low-density suburbs changes the environment physically by eliminating vegetation (tree-cutting) and by increasing dark surfaces like roads.

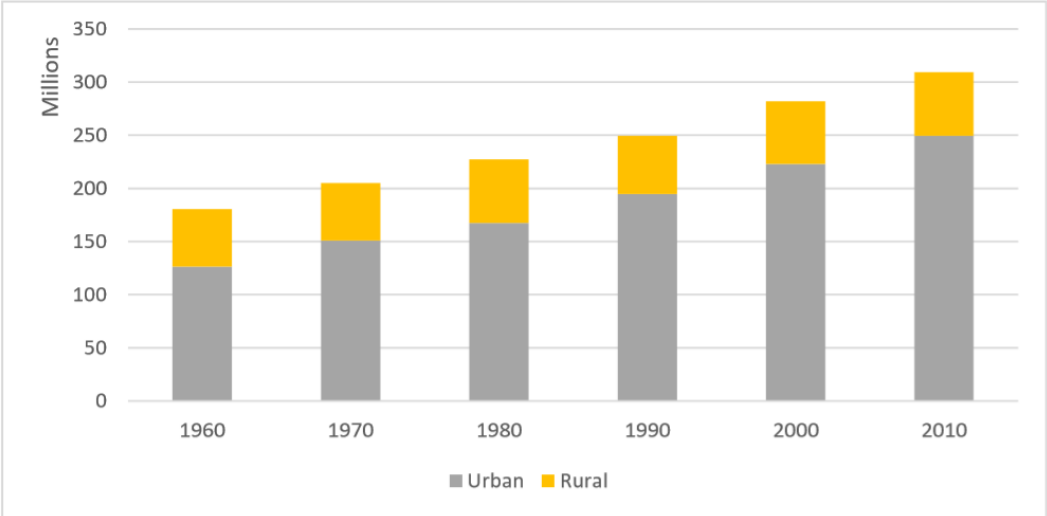


Figure 1.1: Population growth and geographic distribution in the United States
Source: U.S. Census Bureau

The association between urban sprawl and climate has been discussed in the economic, urban planning, and environmental science literatures. However, existing studies have at least one of the following shortcomings. First, and probably the most important, is the lack of a causal study that shows the causal effect of sprawl on climate variables. Not considering the causal relationship limits a study's usefulness for policymaking. The second shortcoming is related to the definition of the two main variables of such studies, namely, sprawl and temperature. Most of the studies in this literature use an overly generous definition of sprawl that prevents one from making a clear statement on the relationship of interest, or

their temperature-related variables are limited. and as a result, they do not examine the various aspects of the phenomenon. In Oke (1973) the relationship between population and the intensity of the UHI effect in Montreal is explored and it is found that areas with higher density are correlated with a grade of the UHI effect. This is not a surprising result, as many of the factors that cause the UHI are stronger in more densely developed areas of a city. From a public policy point of view, the important question is whether to build vertically or horizontally. The answer in the literature depends on how the question is asked because of the interdisciplinary nature of the question and how UHI is measured. Similarly, (Coutts et al. (2007)) using data from multiple sites across Melbourne and (Martilli (2014)) using simulation data, find a contradictory result that compact planning leads to less UHI formation. That is because, in both studies, compact sites contrast with low residential sites instead of comparing compact growth plans versus sprawling growth patterns. In Stone and Rodgers (2001), the contribution of residential development patterns to the development of a surface-heat island in the metropolitan region of Atlanta, Georgia, was explored. Their result suggests that lower density residential patterns lead to a higher grade of UHI formation. In Stone et al. (2010), the correlation between the mean annual change in the number of extreme heat days between 1956 and 2005 and the sprawl index of each region in 2000 is measured, where sprawl is measured as in Ewing et al. (2003). This study suggests that occurrence of extreme heath events has been increased over last fifty years, and the rate of this growth in most sprawling MSAs is around double the rate in most compact MSAs.

To construct a measure of sprawl, I use US conterminous Wall-to-Wall Anthropogenic Land Use Trends (NWALT), created in 2015 by the US Geological Survey (USGS). NWALT is a consistent and long-period independent time-series dataset that depicts land use and covers the United States over five waves between 1974 and 2012. Utilizing NWALT and following Burchfield et al. (2006), I construct panel data of residential compactness, which is the dimension of sprawl that I am interested in. This provides, for the first time, panel data of sprawl that are based on the actual expansion of 350 metropolitan areas.

Climate and, in particular, temperature is one of the factors that affect households' location choice. To address endogeneity concerns, I utilize the number of interstate highways in the national plan, emanating from the central city, as an instrument for sprawl. My finding suggests that, for the middle-sized metropolitan statistic areas (MSAs), a 10 percentage-point decrease in residential compactness leads to about a 1.1 degree Fahrenheit increase in the MSA center's annual mean temperature.

The rest of the paper is organized as follows. Section 2 describes the construction of urban sprawl. In Section 3, I discuss the estimation methods. Section 4 discusses the construction of two different data sources (weather data and residential compactness). Section 5 presents the results. Section 6 introduces a new method for targeting the proper control group and, finally, Section 7 concludes the paper.

1.2 Measuring Urban Sprawl

Urban sprawl refers to a particular form of urbanization that is associated with certain characteristics. Studies' definitions of sprawl differ based on the aspect of the sprawl considered, and limitations in data and methods. However, the following characteristics have been widely associated with urban sprawl. First, low population density and a high level of urbanized land per person, which indicates inefficient land use. When in a given area, the rate of urbanization is much greater than the rate of population growth, we face a sprawling phenomenon (Black (1996); Freeman (2001); Galster et al. (2001), Harvey and Clark (1965), Glaeser and Kahn (2004), Baum-Snow (2007), Ewing et al. (2003)). Second, leapfrogging or scattered development, which refers to the building of new residences, either separately or in a subdivision, at some distance from existing built-up areas, especially in the transition zone between urban and rural areas (Clawson (1962), Mills (1981), Gordon and Richardson (1997), Yeh (2001), Burchfield et al. (2006), Ewing et al. (2003)). Third, separate land use, which is when employment and retail services are a significant distance from residential areas, which increases driving (Brown et al. (1998), Duany et al. (2001), Ewing et al. (1994), Ewing

(1997), Ewing et al. (2003)),Ewing and Hamidi (2014), Galster et al. (2001)). Fourth, lack of street accessibility and connectivity, or unplanned urban growth in the suburbs, which leads to inefficient street systems (Duany et al. (2001), Allen and Benfield (2003), Ewing 1994, 1997, 2003, 2014).

Although the presence of sprawl is obvious, it is difficult to define and quantify. In fact, even though the sprawling city has been a hot topic since the early 1950s, it has not been properly quantified until recently Malpezzi et al. (1999)). In Galster et al. (2001) eight different measures of residential development are reviewed to grasp different dimensions of sprawl, including density centrality, proximity of land use, etc. The study ranks 13 large US cities based on six of these measures. In Ewing et al. (2003), sprawl indices for 83 US metropolitan areas for change between 1990 and 2000 are estimated using 22 variables representing various aspects of development patterns. They focus on four different dimensions of sprawl—residential density, land use mix, degree of centering, and street accessibility—arguing that one factor alone cannot capture the complexity of sprawl. And while some cities like Atlanta sprawl in all dimensions and others like New York are compact in all factors, other cities are not consistent in all factors. To construct a sprawl index for each dimension they combine up to seven variables via principal component analysis into one factor representing the degree of sprawl in each dimension. In Ewing and Hamidi (2014), this result is updated for 221 metropolitan areas in 2010. Unfortunately, most of the variables they use are not available before 1990. Thus, their index cannot be used to capture sprawl trends over time. Moreover, reducing many dimensions to one index involves some degree of information loss. To study the effect of sprawl it is important to have a clear idea of which factor plays a role.

A variety of measures are used in Glaeser and Kahn (2004) to capture sprawl focusing mostly on population density and separation of use. Their measures include percentage of population density and job density (within Inner 3 and 5 Mile Ring and MSA) and median person's distance in miles from the central business district (CBD). They also report that

the correlation between different measures can be very low. In Angel et al. (2005), classifying satellite images of cities in 1990 and 2000 directly examines the expansion forms of cities. The study uses population density as the main variable of interest, but the index of population density has a strong point compared to previous studies. Instead of administrative boundaries, they measure the actual built-up area of the city. Using administrative boundaries does not allow a reliable comparison between cities. So, the resulting population density is sensitive to the definition of boundaries which varies even within the United States. Another weakness in using average population density is that it neglects the distributional aspect of the metropolitan population.

Burchfield et al. (2006) focus on capturing the extent to which residential development in urban areas is scattered. Like Angel et al. (2005) their methodology is based on analyzing landscape change with satellite remote sensing and a geographic information system (GIS). They use land cover and land use data from Landsat 5 Thematic Mapper satellite imagery and high-altitude aerial photographs. Their data contain square cells of 30 x 30 meters situated on a regular grid. Each grid cell is assigned to one specific land use code, such as residential development, water, forest, etc. To measure the extent of sprawl, for each 30-meter cell of residential development, they calculate the percentage of open space in the immediate square kilometer. They compute the sprawl index as a change in the average undeveloped land across all residential development in each metropolitan area. They also introduce the level of development for two periods by calculating the percentage of land not developed in the square kilometer surrounding residential areas for 1976 and 1992. They discuss the correlation of their measure with other measures such as median lot size, miles driven per person, and share of employment over 3 miles from the CBD and conclude that while scatteredness is a key factor of sprawl, it does not grasp all of its dimensions. Specifically, they find a low correlation between their measure and centralization of employment.

Following Burchfield et al. (2006), this study utilizes data obtained from remote sensing to construct a measure for urban sprawl. The benefits of using remote sensing data are

twofold. First, they are consistent over great areas and over time, which enables the construction of panel data with vast geographical coverage. Second, it is possible to measure the expansion of the built-up area of cities directly. I measure residential patterns using land use data and, by taking the average value across all residential cells within a metro region, to construct an index for residential compactness. The first two characteristics of sprawl, namely low population density and scattered development, are often found not to be correlated with the second two characteristics, that is, separate land use and lack of street accessibility. To distinguish the effects of these characteristics, instead of measuring the ratio of open space in each cell's neighborhood, I measure residential area and commercial area separately. I also extend Burchfield et al. (2006) in both geographical and temporal coverage and construct a panel dataset of 363 metropolitan areas for 1974, 1982, 1992, 2002, and 2012. In the following, I explain the construction of residential compactness.

1.2.1 Residential Compactness

I use US conterminous NWALT, which gives a national 60-meter, 19-class mapping of anthropogenic land use for five time periods. NWALT is compiled using existing data sources including NLCD 1992, 2006, and 2011, the USDA Census of Agriculture, 1974–2012, and Spatial Analysis for Conservation and Sustainability 1970–2000. Sprawl measure, which was introduced by Burchfield et al. (2006), is the ratio of undeveloped cells in a neighboring square of 1 square kilometer, centered at a residential cell, averaged over all the residential cells across the MSA. In this study, I use the following procedure that involves assigning the percentage of residential area around each residential cell (with a radius of 560 meter) and then averaging over all the assigned percentages in MSA:

1. For every residential cell $i \in I_k$ where I_k is the set of all the residential cells in MSA_k,

count all the residential cells in a circle with a radius of 1 kilometer around i:

$$Count_i = \sum_{j \in J} 1[j = residential] \quad \forall i \in I_k$$

where $Count_i$ is the total number of residential cells around residential cell i; $1[.]$ is an indicator function that equals one if cell j in the neighborhood of residential cell i is residential and zero otherwise; and J is the set of all cells in a circle (neighborhood) with a radius of 1 km around residential cell i.

2. Divide $Count_i$ by the total number of cells in the neighborhood around cell i:

$$rc_i = \frac{Count_i}{|J|} \quad \forall i \in I_k$$

where rc_i is the residential compactness ratio calculated for cell i, and $|J|$ is the norm of the set J.

3. Averaging all the rc_i's in MSA_k results in the residential compactness index for the MSA_k. To calculate the relevant measures of residential compactness as described, I use ArcGIS software. Construction of the RC_k for the Atlanta metro area is shown in Figure 1.2a through 1.2c.

$$RC_k = \frac{\sum_{i \in I_k} rc_i}{|I_k|}$$

The residential compactness RC_k, is zero if there is an MSA in which for every residential cell I_k, $Count_i$ is zero. In other words, there is no other residential cell in a neighborhood of 1 kilometer radius. In contrast, if there is a MSA completely developed for residential use, rci is one for every $i \in I_k$ and by construct, $RC_k = 1$. This reflects the highest level of residential compactness.

1.2.2 MSA Center

To identify an urban area, I need to define the center of each MSA, assuming there exists only one and unique center around which the city is developed. This follows on from the traditional monocentric urban model first introduced by Alonso (1964) and developed by Wheaton (2006), among others. I set the center of the city as the CBD. Using the same procedure as residential compactness, I introduce the Commercial Accessibility measurement. This measurement calculates the percentage of the commercial area around each residential cell (with a radius of 5 km) and then takes the average over all the residential cells for each MSA. This Measure captures the separate land use or employment accessibility that Both Galster et al. (2001) and Burchfield et al. (2006) find nearly uncorrelated with the residential distribution indexes which are designed to measure the sprawl. The procedure follows:

1. For every residential cell $i \in I_k$ where I_k is the set of all the residential cells in MSA_k, count all the residential cells in a circle with a radius of 5 kilometer around i:

$$Count_i = \sum_{j \in J} 1[j = commercial] \quad \forall i \in I_k$$

   where $Count_i$ is the total number of residential cells around residential cell i; $1[.]$ is an indicator function that equals one if cell j in the neighborhood of residential cell i is residential and zero otherwise; and J is the set of all cells in a circle (neighborhood) with a radius of 5 km around residential cell i.

2. Divide $Count_i$ by the total number of cells in the neighborhood around cell i(radius of 5 km):

$$ca_i = \frac{Count_i}{|J|} \quad \forall i \in I_k$$

   where, ca_i is the commercial accessibility ratio calculated for cell i, and $|J|$ is the norm of the set J.

3. Averaging all the rc_i's in MSA_k constructs the commercial accessibility measure for the MSA_k.

$$CA_k = \frac{\sum_{i \in I_k} rc_i}{|I_k|}$$

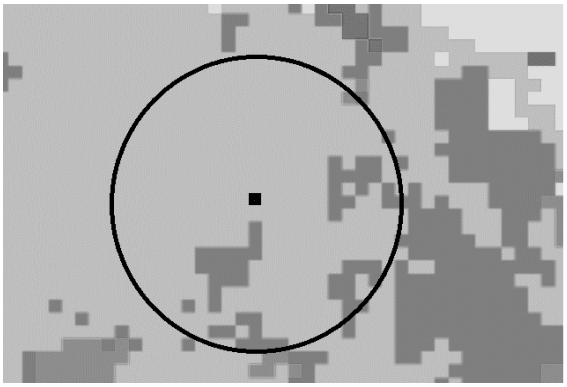
To calculate the relevant measures of commercial accessibility as described, I use ArcGIS software. Construction of the CA_k for Atlanta Metro area is shown visually in Figure 1.2d

I define CBD as a point with the highest commercial accessibility (CA_k) ratio in each MSA_k. The position of the MSA center plays an important role in my analysis because it determines which weather stations belong to the inner city area and which weather stations can be considered outer city. Figure 1.3a and 1.3b show how construction of inner city and outer city regions takes place.

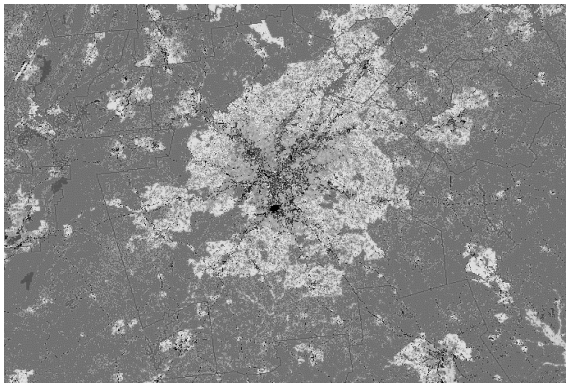
1.3 Estimation

1.3.1 Identification Strategy

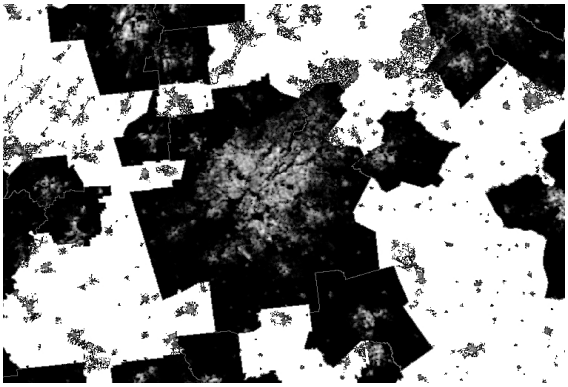
To identify the effect of horizontal urban development on inner city temperature, one has to take into account the reverse causality problem that emerges when temperature change in the inner city or suburb affects an individual's decision to relocate in or out of the city center. To detect the causal effect instead of the correlation structure I use the development of highways in the United States as an exogenous intervention or policy change. In their land use model, Alonso (1964) suggest households consider a combination of rent and transportation costs as the price of housing. Hence the development of highways should shape new suburbs as households decide to relocate and optimize their utility. However, the development of highways can be used as an intervention that affects residential compactness, only if it is exogenous. Following Baum-Snow (2007) and Duranton et al. (2014), it can be argued that since the development of new highways was based on national defense and trade needs,



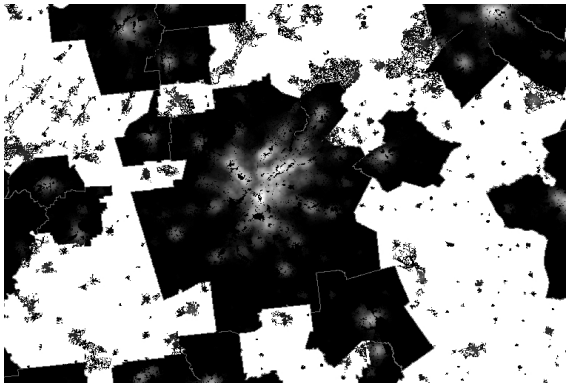
(a) NWALT Data (60m × 60m cells)



(b) Metro Area



(c) Residential density pattern

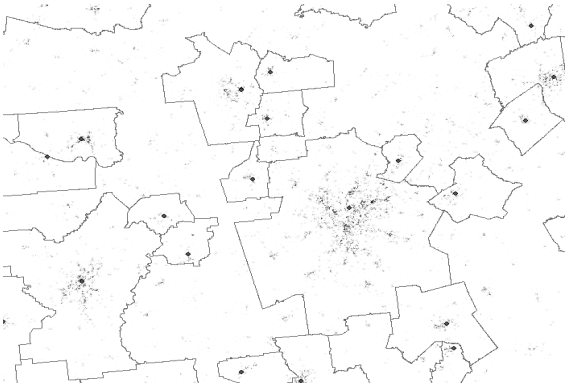


(d) Commercial density pattern

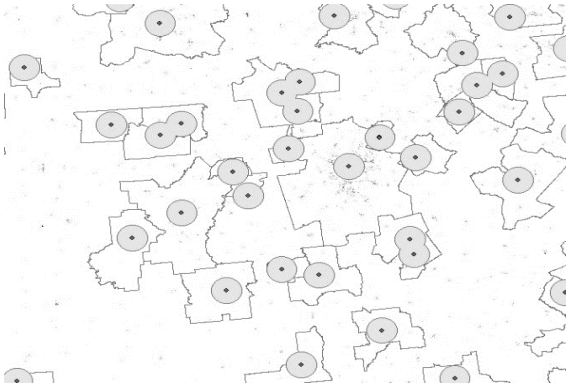
Figure 1.2: Construction of Residential Compactness and Commercial Accessibility ratios for Atlanta metro area

Source: Author's Calculation using NWALT Data and ArcGIS

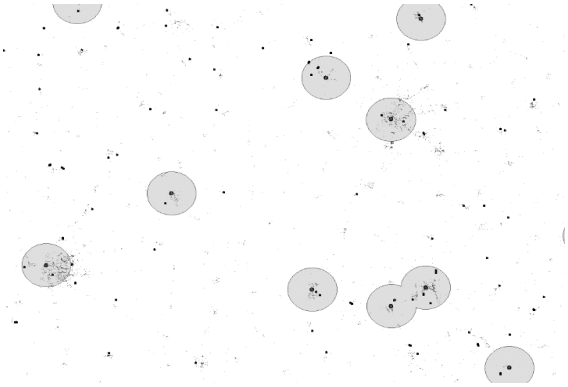
National 60-meter, 19-class mapping of anthropogenic land uses, with each 60m × 60m cell have a value that corresponds to a certain classification. In particular it allows to identify residential and commercial land use. Shown in each figure is an enlarged section of the raster data. More information on classes and subclasses can be find in Falcone (2015). (a) For each residential cell flagged by the land use data (black cell in center of circle in figure), I divide number of residential cells in the neighborhood within 560 m radius by the total number of cells in the neighborhood to get the percentage of residential cells. This ratio then is assigned to each residential cell. (b) This Figure shows Atlanta Metropolitan area featuring 19 class land use. Metropolitan Statistical Area (MSA) lines are depicted for the MSA by merging MSA vector data and NWALT. (c) This figure shows the residential ratio in neighborhoods of each residential cell. Lighter areas show more compact residential neighborhood relative to darker ones. (d) This figure shows spatial distribution of the commercial accessibility in Atlanta MSA. For each residential cell flagged by the land use data, I have count number of commercial cells in the neighborhood with 5 km radius, and divide it by the total number of cells in the neighborhood to get the percentage of commercial cells in neighborhood. This ratio then is assigned to the each residential cell. The Lighter areas show higher commercial accessibility relative to darker ones.



(a) Central business district



(b) City bound



(c) Locating the weather stations on the map

Figure 1.3: Finding Central Business District (CBD) and setting the city bound

Source: Author's Calculation using NWALT Data and ArcGIS

(a) I define CBD as the center of the most concentrated circle with 5 km radius in each MSA (residential cell with highest commercial accessibility). This provides me with a monocentric city within each MSA. (b) This figure shows a ring around each CBD as the inner city. (c) Combining the map of the inner city bounds with the weather station coordinates, and utilizing the data from the vector MSA, NWALT and GSOD, I can locate the stations, in the neighborhood of each MSA center.

requiring the shortest possible distance to connect MSA centers, its map is not based on household preferences.

1.3.1.1 Trend Analysis

The first step to evaluate the temperature change in urban and non-urban area is to establish if there is a trend. One approach is DiD, which is developed from the randomized controlled trial (RCT) literature. The first interstate highway plan was completed in 1974 and the last in 1992. The remote sensing data go back to as early as 1972 (wave 1) and temperature data are sparse for the years before 1970. Hence, I use 1970 as the first period. Since development of highways continuously changes, I do not need to use pre/post periods. In fact, one can compare any two points in time as different levels of development of highways provide continuous DiD [3]. I use Wave 4 (years 2000 to 2004) as the final period using this approach. This allows households to relocate after completion of highways in Wave 3 (1990–1994). I compare outcomes in each city center to a region close enough to MSA center so that it reflects the MSA's characteristics but is not affected by the development of highways and urban development. consequently, I estimate the following regression model:

$$Y_{it} = \beta_0 + \beta_1 Post_{it} + \beta_2 Center_{it} + \beta_3 Post_{it} Center_{it} + X_{it}\beta_4 + \epsilon_{it} \tag{1.1}$$

where Y_{it} is the outcome for MSA i at period t (1972-1976, and 2000 - 2006), $Post_{it}$ refers to a dummy indicating one if it is post-intervention and zero otherwise; $Center_{it}$, determines if the observation belongs to MSA center or the control group (MSA surrounding area). The DiD coefficient is β_3 which compares trends in the city center relative to outskirts. I control for other characteristics by X_{it} which is a vector of time-varying, MSA-specific variables. Lastly, ϵ_{it} is a gaussian error that enables the utilization of simple OLS for this regression.

[3] This approach has been utilized by Acemoglu et al. (2004) among others.

1.3.1.2 Instrumental Variable

As discussed in the previous section, highways are likely to affect sprawl as a consequence of decreased transportation costs. This insight is based on the theory, developed by Alonso (1964), for land use. If transportation costs decrease, then demand for space in the suburbs relative to central cities increases. In Baum-Snow (2007) the hypothesis that highways contribute to growth in the suburbs as opposed to the central city is empirically tested. As noted by Baum-Snow (2007), in testing such a hypothesis, one might be concerned about reverse causation if urban patterns affect the location of highways. To address this concern, Baum-Snow (2007) and Duranton et al. (2014) use the national plan of highway routes proposed in 1947 as an instrument for highway rays. Since the planned portion of the interstate highways was required to serve national defense and trade goals, the number of rays in the 1947 national plan is a valid source of exogenous variation in highways. According to a press release issued by the Public Roads Administration in 1947, the interstate plan was designed to connect principal metropolitan areas, cities, and industrial centers by routes as direct as practicable to serve the national defense and to connect suitable border points. The authors of Baum-Snow (2007) used the highway plan as an instrument for highways and found a causal connection between highways and sprawl. Continuing along this line of reasoning, as the planned portion of highways affects sprawl and the scatteredness of residential distribution, and residential distribution affects temperature, I test the null hypothesis of no causal effect of sprawl on the temperature of the MSA's center by instrumenting for residential distribution. Hence, cutting out the middle step, I use the planned portion of highways as an instrument for sprawl in causal analysis of the effect of sprawl on UHI. I calculate the time series of rays by multiplying the number of rays in the 1947 plan and the fraction of federally funded highway mileage in the 1956 Federal Aid Highway Act completed at each point in time. For the first stage in my two-stage IV, I have:

$$RC_{it} = \alpha + \beta PlannedHW_{it} + \theta Z_{it} + d_i + \epsilon_{it}, \tag{1.2}$$

where RC_{it} is residential compactness; Z_i is the vector of control variables; d_i is MSA-specific fixed effects that point out the utilization of panel data; and $PlannedHW_{it}$ is the planned portion of interstate highways. Note that I use planned portion of highways instead of completed portion following Baum-Snow (2007), to control for reverse causality between completed highways and residential distribution through MSA as it may affect completion speed and allocation of funds. After the first stage, I use predicted values of the measure of residential compactness in the second stage:

$$Y_{it} = \delta + \theta \hat{RC}_{it} + \lambda Z_{it} + e_{it}, \tag{1.3}$$

where Y_{it} represents an environmental variable (temperature for the main analysis) such as the annual mean temperature of the MSA center.

1.3.1.3 UHI Model

There has been much emphasis on observing the level and change in the magnitude of UHI and factors that contributed to this phenomenon trough time. In the previous section, I discussed the Impact of Urban sprawl on urban and rural areas separately. To examine the Impact of Sprawl on UHI directly, I integrate the IV and DiD strategies. I assume that the temperature of the surrounding areas of the MSA center is a control for the temperature of the inner city for each MSA. Since the distance between the inner and surrounding areas is relatively short, the control group can reflect the same unobserved heterogeneities of the inner city area. For the second difference, I exploit the exogenous variations by my IV, planned highway's ray. Assuming that the planned portion of highways exogenously affects residential compactness, I have a policy change (exogenous variation) for each MSA-year and hence can employ the continuous DiD identification strategy for every wave in my panel (instead of only two waves in continuous DiD). The combination of IV and continuous DiD adds to the identification power by the factor of the number of MSAs, multiplied by the number of waves. This is due to the fact that each MSA receives different doze of treatment,

which increases the verifiable domain of the treatment (instead of only binary limits of no treatment and full treatment). However, to make sure the two identification strategies (IV and continuous DiD) work in accordance with each other, exogenous variation from IV should not affect my control groups (temperature of surrounding areas). The combination of IV and DiD also allows me to exploit the panel data features to control for individual fixed effects. The MSA fixed effect can provide some degree of identification power due to the fact that it reduces the chance of unobserved heterogeneity bias. In other words, if there is a fixed characteristic associated with an observation (e.g., elevation), I do not need to control for it since it is reflected in MSA fixed effect. Other characteristics that are variable through time can be controlled by the continuous DiD setup, where I have a control observation for each inner-city point in time. The first stage of the two-stage panel IV is the same as the first stage of the two-stage IV, which was explained previously:

$$RC_{it} = \alpha + \beta PlannedHW_{it} + \theta Z_i + d_i + \epsilon_{it}, \tag{1.4}$$

where the definitions of the variables are the same as they are in equation (2). In the second stage, instead of using the temperature of the inner city or the MSA's center, I use the difference between the temperature of the MSA's center and its surrounding area, or the treated and control group. The definition of these groups is the same as that explained in the continuous DiD section. Thus, the second stage of the two-stage IV is:

$$\Delta(Y_{it}) = \delta + \theta \hat{RC}_{it} + \lambda Z_i + e_{it}, \tag{1.5}$$

Here, Δ refers to the difference between interested outcome in and out of the MSA's center. Since the main focus of this study is to address the UHI effect, I continue using temperature as the outcome Y. The setup here allows for the panel regression techniques such as first difference and fixed-effect, where both of them help to control for individual unobserved heterogeneities.

1.4 Data

1.4.1 Temperature Data

I use Global Summary of the Day (GSOD) data for years 1972–1976, 1980–1984, 1990–1994, 2000–2004, and 2010–2014 (25 years) to calculate the desired statistics such as annual and seasonal average temperature. GSOD data provide 18 surface meteorological elements which are driven using hourly observations from USAF DATSAV3 surface data and federal climate complex Integrated Surface Data (ISD). Historical data are available from 1929, but the quality of the data are better after 1973, in terms of the number of stations and the average number of reported days per station. To make a daily observation, a minimum of four observations are needed for the day. Thus, for station-days with less than four observations, GSOD reports are missing. Other causes of missing observation are data restrictions or communication problems. For this study, I use only one year (1972) from the earlier part of the GSOD with less quality. For the year 1972, I only observe 53 stations, in contrast to an average of 394 stations per year in later years. The total number of weather observatory stations I observe is not constant through time because a station can stop working for some years. I include a station-year in my data if it occurs inside a MSA boundary for that particular year. The number of stations included in the data and their changes are summarized in Table 1.1. For years after 2010, the number of stations increases from 394 to 900 on average. Also, for every station-year that is available, weather information for 340 days is reported on average. I use all the available data, which total 4,228,407 station-days. Then for each station, I compute statistics such as annual and seasonal mean, standard error, and maximum and minimum temperature.

1.4.2 Assigning Temperature to MSA Center

As mentioned before, temperature data are station-year specific, and I need a method to assign the acquired temperature from a station to a particular MSA center. Taking the

Table 1.1: Availability of the weather stations in GSOD data

Year	Total Stations	Total Station-Days	Average Days per Station
1972	53	19,063	359.68
1973	380	131,430	345.87
1974	381	132,380	347.45
1975	406	135,461	333.65
1976	417	141,561	339.47
1980	448	148,988	332.56
1981	448	147,603	329.47
1982	445	145,624	327.24
1983	452	147,706	326.78
1984	453	149,142	329.23
1990	464	152,433	328.52
1991	448	148,905	332.38
1992	448	147,864	330.05
1993	444	148,687	334.88
1994	434	146,490	337.53
2000	322	113,680	353.04
2001	330	114,075	345.68
2002	339	116,707	344.27
2003	341	118,020	346.1
2004	417	142,249	341.12
2010	942	325,056	345.07
2011	911	320,713	352.05
2012	904	320,319	354.34
2013	881	307,766	349.34
2014	863	306,485	355.14

Number of weather stations, available in the data and their changes through the study period.

MSA center's geographic coordinates (latitude and longitude) by the construction explained in Section 2, and coordinates of the stations from GSOD data, I can calculate the distance between every station/center pair. For this purpose, I use a planar approximation and limit the distances to less than 250 kilometers. Two deficiencies of planar approximation are where approximation is not exact. This occurs first when the distance between two points is very long and, second, when one or both points are close to the geographic pole. I do not have the second problem since the northernmost latitude in the contiguous United

States is 49.38407°N. Also, limiting the search area to a circle of 250 kilometers solves the first problem. I project the latitude and longitude coordinates onto a plane assuming the spherical earth, using the following formula:

$$D_{ijt} = \sqrt{((\cos \frac{\pi}{180} y_{it}) \times 111.321 \times \Delta x_{ijt})^2 + (\Delta y_{ijt})^2} \tag{1.6}$$

where x_i t and x_j t are the longitude of MSA center i, station center j, in period t, in degrees, and y_i t and $y_j t$ are latitudes in degrees. Then, the distance between MSA center i and station j is calculated in kilometers ($D_i jt$). Also, the formula corrects for the variation in distance between meridians (longitudes) with latitude. This problem occurs when the distance between two points on two different longitudes and on the same latitude shrinks as we get closer to one of the poles and further from the equator. This formula also helps us to avoid the computational burden caused by assuming a non-spherical earth in the other formulae. With the distance in kilometers between the MSA center and all stations in a radius of 250 kilometers, I calculate the average temperature of stations that are in a circle with a particular radius r, centered at the MSA center:

$$IB\bar{T}_{it|r} = \frac{\sum_{j \in J_{i|r}} c_{jt}}{|J_{i|r}|} \tag{1.7}$$

where $IB\bar{T}_{it|r}$ is the average temperature assigned to the MSA center i in period t, using radius r, and is called inner bound temperature (IBT); $c_j t$ is the seasonal or annual statistics related to temperature (mean, max, min) in station j and time t; J_{ir} is the set of all stations in the radius r of MSA center i and $|J_{ir}|$ is the norm of the set J_{ir}.

To analyze the UHI effect, I need to assign another temperature to the area around the MSA center that represents the suburb of the center. I call this second temperature the outer bound temperature (OBT). I calculate the OBT using:

$$OB\bar{T}_{it|rR} = \frac{\sum_{j \in J_{i|rR}} c_{jt}}{|J_{i|rR}|} \tag{1.8}$$

where $OB\bar{T}_{it|rR}$ is the mean temperature of the stations in the surrounding areas of MSA center i,at time t, using an inner radius of r and outer radius R; $J_{i|rR}$ is the set of all stations located in distance between r and R to the MSA center; $c_j t$ is the temperature of center j at time t and $|J_{i|rR}|$ is the norm of the set $J_{i,rR}$. Figure 1.3c shows the location of weather station relative to metropolitan area for the data sample.

1.4.3 Setting Bounds

The MSA is a geographic area with high population concentrated at its core and surrounding areas with economic ties to the core. A MSA requires a Census Bureau urbanized area of at least 50,000 population. This definition can result in an enormous heterogeneity among MSAs. Furthermore, one particular MSA might grow or diminish in size through time. As a result, using only one fixed inner radius r and outer radius R for all MSAs might be problematic due to the urbanized area differences between small and big MSAs and through time for one specific MSA. I approach this problem by introducing flexible bounds that involve assigning the radius $r_i t$ and R_{it} to different MSAs in different periods. I define both radii r_{it} and R_{it} as a function of the total number of residential land cells ($TRLC_{it}$) in each MSA i at time t:

$$r_{it} = \frac{\sqrt{\frac{TRLC_{it}}{\pi}}}{\frac{\sqrt{\frac{TRLC_F}{\pi}}}{r_F}} \quad , \quad R_{it} = \frac{\sqrt{\frac{TRLC_{it}}{\pi}}}{\frac{\sqrt{\frac{TRLC_F}{\pi}}}{R_F}} \tag{1.9}$$

The formula considers the fact that since $TRLC$ is the area of residential land, the radius can be calculated from $TRLC$ using the formula for the circle area ($Area = \pi \times r^2$). Then this radius, which represents the numerator of both formulae above, is adjusted by the scaling parameter in the denominator:

$$\frac{\sqrt{\frac{TRLC_F}{\pi}}}{r_F} \quad , \quad \frac{\sqrt{\frac{TRLC_F}{\pi}}}{R_F} \tag{1.10}$$

Both are calibrated from the average of the LC_{it} , r_{it} and R_{it} of the MSAs with distinctive inner and outer radiuses. I calibrate them using the numbers, as shown in Table 1.2. While the numerator of the functions allows for variations in bound between MSAs with varied sizes, the denominator scales the bound to an agreeable size.

For the fixed bound scheme, I use the radius suggested by the flexible scheme formula

Table 1.2: Calibrated values to being used in flexible bound scheme

$TRLC_F$	r_F	R_F
800000	50	150

for the first period that MSA is observed and then keep the radius constant through time. This allows me to reduce the heterogeneity between MSAs, but changes through time are not reflected by this fixed measure.

1.5 Estimation Results

In this section, I discuss the results for both schemes: fixed and flexible bounds. I follow each with a sensitivity analysis.

1.5.1 Fixed and Flexible Bounds

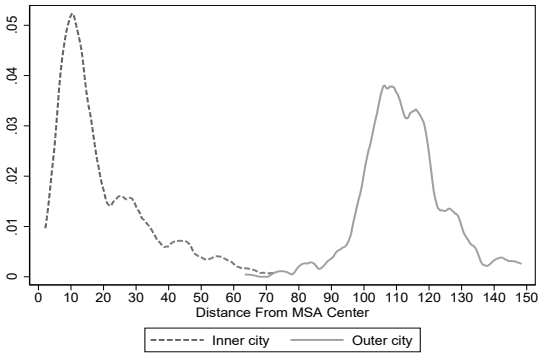
The first step to examine the change in temperature or any other attributes of the urban areas and the role of sprawl in it is to define the Urban area and the rural surroundings carefully. As explained in section 1.4.3, this study employs bounds to distinguish between urban areas (in-bound) and surrounding rural areas (out-bound). Each bound's radius is a function of the MSA's residential counts. In a fixed bound scheme, this residential count is fixed at the level when the MSA is observed for the first time in the data (1972). Whereas, the flexible bounds are updated by residential count of MSAs in each wave.

Fixed bounds provide more intuitive interpretation of the results as geographical boundries of each MSA are fixed through time. On the other hand, flexible bounds are more responsive

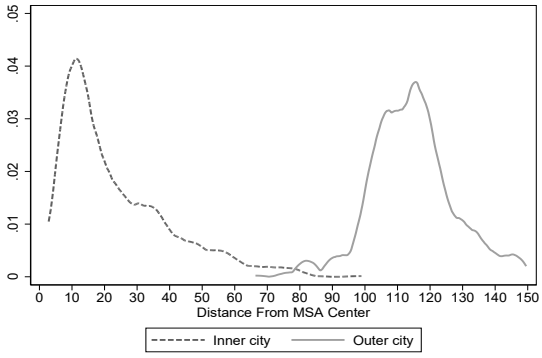
to the changes in urban structure and hence, better reflect the concept of the urban areas and cities. That is, when residential areas spread from the edges of cities, previously rural (surrounding) areas become part of the urban area. In the flexible bound scheme, we address this issue by adjusting the city boundaries. However, while the flexible bound scheme provides a more stable and reliable distinction between rural area and it's surroundings, interpretation of the results produced using this scheme is more difficult. Note that in the flexible bound scheme, urban area is not attached to a specific geographical area, and both in-bound and out-bound regions may expand or shrink in time to reflect the city boundries.

In Figures 1.4a and 1.4b distribution of distance from MSA center to the centroid of the weather stations are shown for fixed and flexible schemes respectively. Comparing two schemes, the distribution of in-bound and out-bound intersect in flexible scheme, while they are relatively separated in fixed scheme. This is due to the growth in some of the MSAs that increases their in-bound radius in the flexible scheme. This stretch in distribution is more apparent when comparing Figures 1.4c and 1.4d. Figure 1.4e and Figure 1.4f show that even though a flexible scheme allows for stretching of the bounds, the overall shape of the distribution of the distance between in bound and out bound is preserved, except for the far right tail of the distribution. Finally, to achieve a better grasp of the changes through time, Figure 1.5a provides growth of the inner-bound radius for a selected group of major cities/MSAs. While they all grow over time, New York grows the slowest and Atlanta grows the fastest. This is partially due to the fact that New York was already developed both horizontally and vertically and, hence, there is not much extra space for horizontal development.

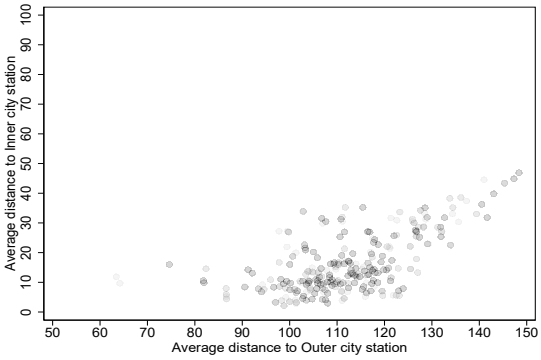
Since we are interested in determining the effect of residential compactness on the UHI effect, it is vital to our analysis that there are enough stations in both, that is, in and outside the bound. The number of MSAs with available data temperature outside the bound is 7842, and those with available data for temperature inside the bound is up to 7234 observations in 25 years. The number of MSAs with both inside and outside available temperature data



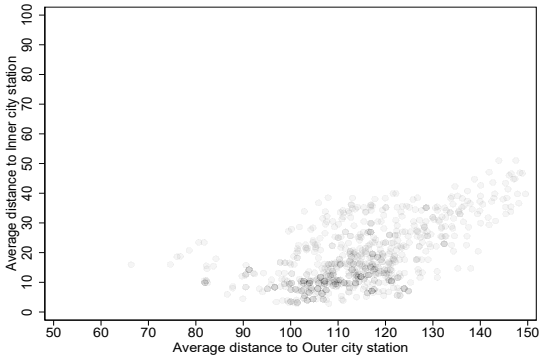
(a) In/Out distribution (fixed bound)



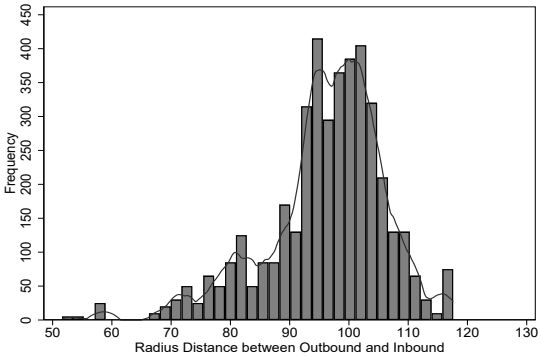
(b) In/Out distribution (flexible bound)



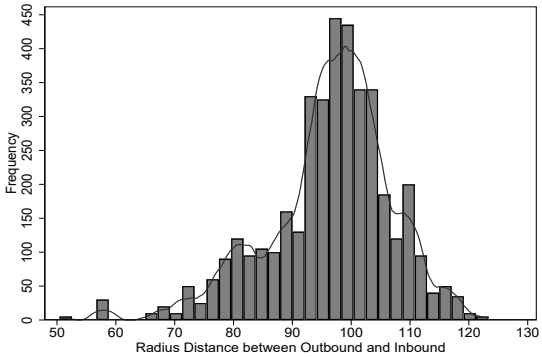
(c) In/Out correlation (fixed bound)



(d) In/Out correlation (flexible bound)



(e) In/Out distance (fixed bound)



(f) In/Out distance (flexible bound)

Figure 1.4: Average distance from MSA center to inner city and outer city weather station

Source: author's calculation using NWALT data and GSOD

(a-b) Distribution of the average distance from MSA center to the inner city and outer city stations, for all the MSA-years. Dashed line shows the distribution of the average distance from MSA center to the weather stations located inside the city bound and solid line is the distribution for the stations located in outer city bound. (a) Fixed bound. (b) Flexible bound

(c-d) The scatter plot of all the MSA-yeras, where horizontal axis measures the distance from MSA center to the outer city stations and vertical axis shows the distance from MSA to inner city stations (c) Using fixed bound, positive correlation pattern appears as expected. It is due to (d) Flexible bound incorporates more stations and MSA-years into analysis.

(e-f) Distribution of the average difference between the inner city and outer city weather station. (e) Fixed scheme. (f) Flexible Scheme.

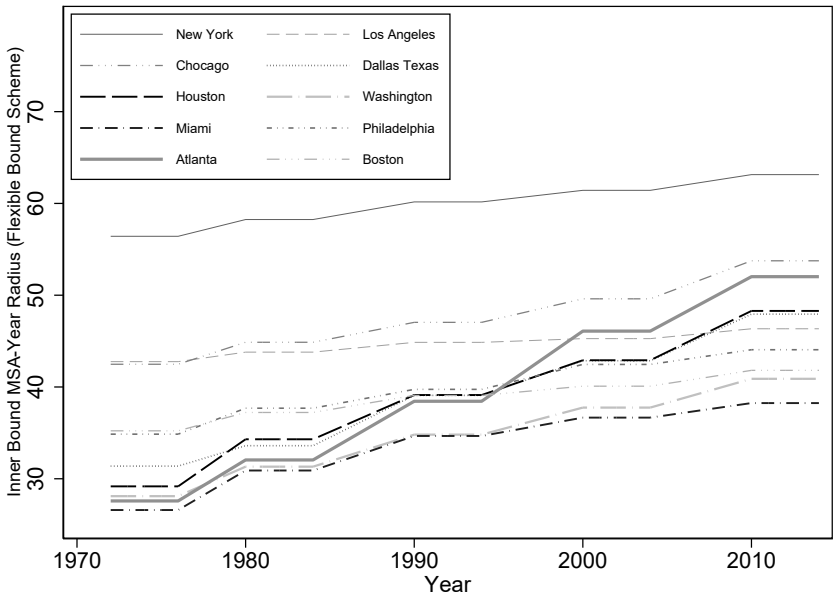
sums to 6715, and over 25 years, it is approximately 268 MSAs per year. This number will then be reduced when we incorporate the IV since IV is not available for all the MSAs. Table 1.3 provides the average distance between existing stations outside and inside the 50 kilometer bound and CBD for all years.

Table 1.3: Number of MSA's and average distance between center and In- and out-bound stations

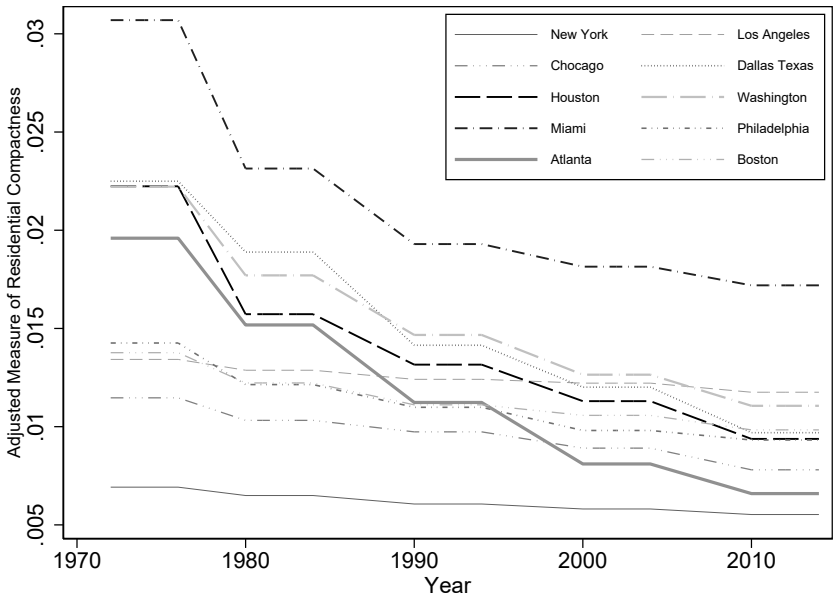
	Number of MSAs	Average distance
Inside the bound	7,234	24.98
Outside the bound	7,842	106.15

1.5.2 OLS Estimates

I use OLS estimator to identify the direction of the reverse causality later when I introduce unbiased estimators. Table 2.16 shows that the increase in residential compactness from zero to one (full-range jump in the RC ratio) reduces the annual temperature of the MSA center by about 10 degrees Fahrenheit annually. This decrease in temperature is more severe during winter and autumn. Residential compactness is moving against residential scatteredness or sprawl and hence, the negative correlation is in accordance with the UHI hypothesis. OLS is a biased estimator, suffering from different types of endogeneities. Reverse causality bias channels the effects from the temperature of the MSA center to residential compactness by affecting household preferences and hence, their decision to relocate to the suburbs or the city center. While factors such as gasoline price may discourage households from moving away from business centers (MSA center), other factors such as avoiding crowded places and possible lower utility prices may encourage them to relocate to the suburbs.



(a) Changes in the Radius size of inner bound



(b) Residential Compactness for selected cities

Figure 1.5: MSA size and Residential Compactness

Source: author's calculation using NWALT data and ArcGIS

(a) Time series of the inner radius, associated with the selected big cities in U.S. from 1972 to 2014, using flexible bound. Atlanta shows highest increase in the inner radius, reflecting the fast rate of transformation in the type of land use, to residential use, and growth in the city limits. (b) Changes in the measure of residential compactness for sample of MSAs for the big cities in the U.S. Atlanta experiences the most radical drop in the residential compactness. Meaning that sprawl/scatteredness rate has been increased in Atlanta MSA.

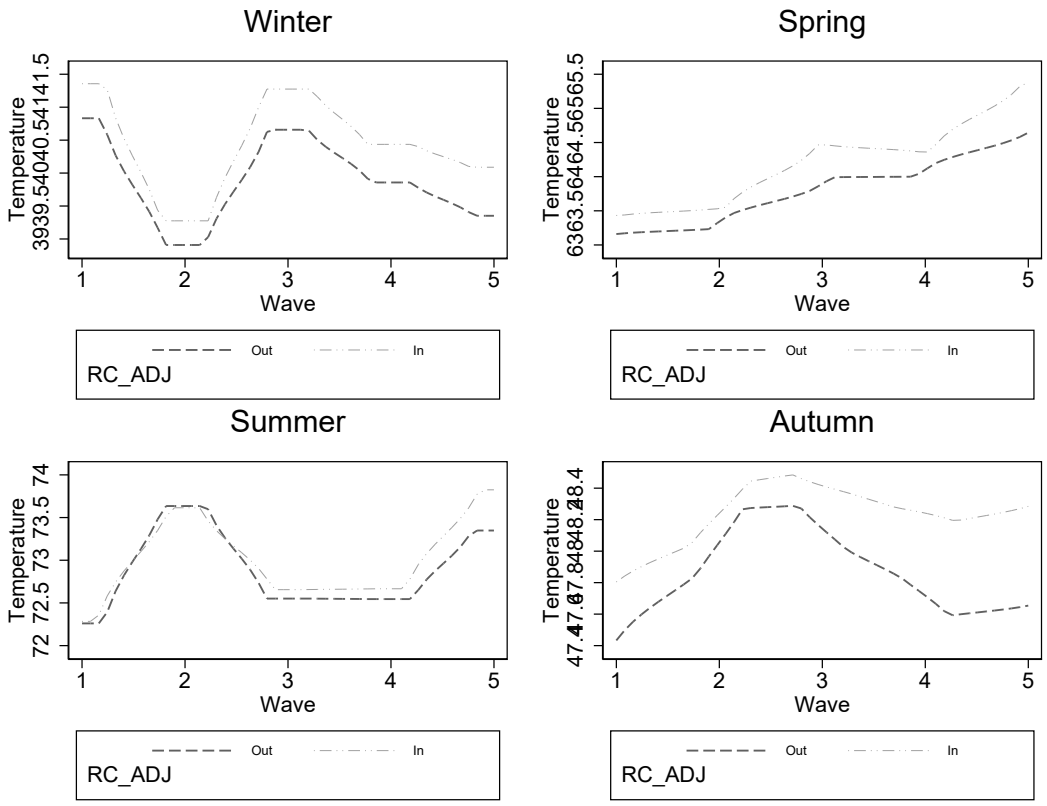


Figure 1.6: Movements of the temperature in time (Fixed bounds)

Source: author's calculation using NWALT data and ArcGIS

Seasonal average of the daily mean temperature is shown for inner and outer city bounds. Temperature of the inner city bounds are on average higher than outer city bounds. The irregularities are due to the fixed bound scheme that does not allow for change in bounds and hence, MSAs which are developing horizontally add to the outer bound temperature and make the inner city and outer city temperatures more similar.

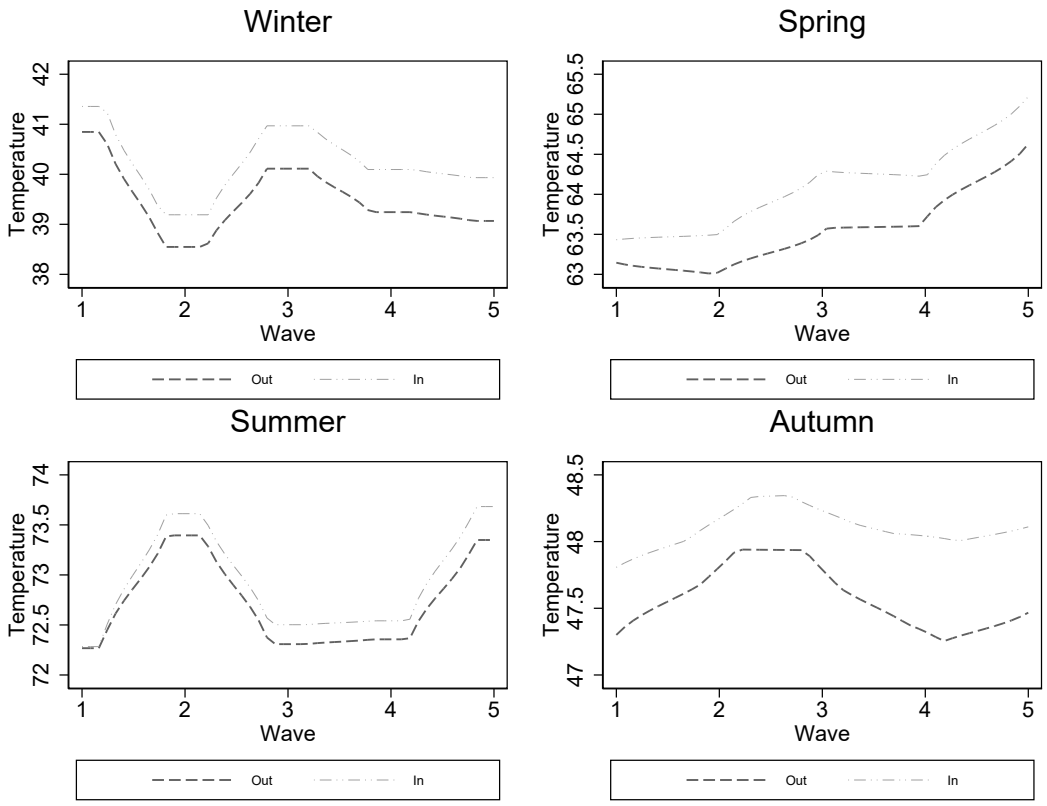


Figure 1.7: Movements of the temperature in time (Flexible bounds)

Source: author's aalculation using NWALT data and ArcGIS

Seasonal average of the daily mean temperature is shown for inner and outer city bounds. Temperature of the inner city bounds are on average higher than outer city bounds. Since, flexible scheme is used, the results are more consistent with the expectation and temperatures in and out of the city bounds are consistently different.

Table 1.4: OLS results of the effects on mean Temperature

(a) Fixed bounds

	(1) TEMP(Annual)	(2) TEMP(Winter)	(3) TEMP(Spring)	(4) TEMP(Summer)	(5) TEMP(Autumn)
RC	-9.771***	-10.749***	-9.526***	-9.306***	-11.134***
	(1.36)	(1.92)	(1.21)	(1.07)	(1.50)
Wind Speed	-0.055***	-0.088***	-0.041**	-0.030**	-0.067***
	(0.02)	(0.03)	(0.02)	(0.01)	(0.02)
Population	0.478**	1.327***	-0.375*	-0.466**	0.984***
	(0.24)	(0.33)	(0.21)	(0.18)	(0.26)
N	3,063	3,100	3,100	3,099	3,097

(b) Flexible bounds

	(1) TEMP(Annual)	(2) TEMP(Winter)	(3) TEMP(Spring)	(4) TEMP(Summer)	(5) TEMP(Autumn)
RC	-9.064***	-11.398***	-8.554***	-8.727***	-11.465***
	(1.36)	(1.89)	(1.19)	(1.06)	(1.47)
Wind Speed	-0.034*	-0.087***	-0.003	-0.002	-0.056***
	(0.02)	(0.03)	(0.02)	(0.02)	(0.02)
Population	-0.108	0.165	-0.643***	-0.577***	0.177
	(0.23)	(0.31)	(0.19)	(0.17)	(0.24)
N	2,921	2,994	2,994	2,993	2,992

Standard errors in parentheses
$*p < 0.1, **p < 0.05, ***p < 0.01.$

1.5.3 Correction for Unobserved Fixed Heterogeneity

One important source of endogeneity is unobserved heterogeneity. This problem can be reduced considerably by employing panel data and estimation techniques that drop fixed individual heterogeneity in the process of estimation and reduce unobserved heterogeneity bias to the individual, time-varying factors that are not orthogonal to the error term. I utilize the fixed-effects panel estimation method that assumes Gaussian error structure and cancels out individual fixed effects. Table 1.5a shows the same results as in the previous section and instead of simple OLS coefficients, it reports fixed-effects coefficients. The overall annual

effect is less negative compared to the OLS estimation. This suggests that MSA-specific features such as elevation could explain a considerable portion of the observed correlation between residential compactness and the temperature of the MSA center, and controlling for them increases the effects, in particular, seasonal coefficients.

Table 1.5: FE OLS results of the effects on mean Temperature

(a) Fixed bounds

	(1) TEMP(Annual)	(2) TEMP(Winter)	(3) TEMP(Spring)	(4) TEMP(Summer)	(5) TEMP(Autumn)
RC	-4.705***	6.331***	-13.388***	-6.805***	-3.059*
	(1.24)	(1.72)	(1.09)	(1.47)	(1.57)
Wind Speed	0.030***	0.012	0.032***	0.023**	0.021**
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Population	-0.215	-0.285	-0.257	-0.313	-0.057
	(0.19)	(0.25)	(0.16)	(0.21)	(0.23)
N	3,063	3,100	3,100	3,099	3,097

(b) Flexible bounds

	(1) TEMP(Annual)	(2) TEMP(Winter)	(3) TEMP(Spring)	(4) TEMP(Summer)	(5) TEMP(Autumn)
RC	-3.495***	6.176***	-11.765***	-5.512***	-2.770*
	(1.23)	(1.67)	(1.06)	(1.44)	(1.56)
Wind Speed	0.019***	0.005	0.019***	0.019**	0.009
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Population	-0.191	-0.376	-0.310*	-0.322	-0.072
	(0.20)	(0.26)	(0.17)	(0.23)	(0.24)
N	2,921	2,994	2,994	2,993	2,992

Standard errors in parentheses
$*p < 0.1, **p < 0.05, ***p < 0.01.$

1.5.4 Trend Analysis Results

Focusing on the reverse causality bias, the DiD estimator is appealing. However, it does not account for unobserved heterogeneity and requires further investigation. Table 1.6a shows the DiD estimates for the fixed bound scheme. At first glance, it is noticable that the

DiD coefficients are all positive (unlike the previous RC coefficients). This is due to the definition of DiD and what it encapsulates. DiD setup here compares in-bound and out-bound averages for the pre- and post-intervention periods while taking into account some observed heterogeneities. Intervention here is a development of US federal highways, which stimulate sprawling. Since sprawl and residential compactness move in opposite directions, a positive DiD coefficient is consistent with the previous results. Also, note that the coefficients for winter and autumn are closer to zero but not negative. The annual MSA temperature is positively and statistically significantly affected by the growth in sprawl or the reduction in residential compactness. Comparing DiD with fixed-effects estimates, it is likely that different biases are not completely opposing each other, and their combination may exacerbate the effects. Unconditional average effects are shown for the interaction of the control/treated and the pre/post groups in Figures 1.8a and 1.8b, which respectively show the results for fixed and flexible schemes. As is apparent, while the average of the treated distribution moves to the right after intervention, for the control group this movement is slightly to the left and hence the overall DiD is positive. Also noticeable is the bi-modal distribution of each group that suggests non-linearity in the effects and possible improvement in DiD precision if two different DiDs are calculated for warmer and colder MSAs. The flexible bound scheme increases the unconditional DiD effect but is very close in shape to the distributions in fixed bound scheme.

1.5.5 The Impact of Sprawl on Surrounding Urban Area Using Instrumental Variable

With panel data, the IV method can in theory address both reverse causality and unobserved heterogeneity concerns. Borrowing the predicted values of highways from Baum-Snow (2007), and using them as an instrument for the constructed RC (changes in the opposite direction of sprawl), I address the reverse causality problem. Furthermore, having longitudinal data, I control for the MSAs' fixed effects and address the endogeneity issue due to

Table 1.6: Diff-in-Diff effects on mean Temperature

(a) Fixed bounds

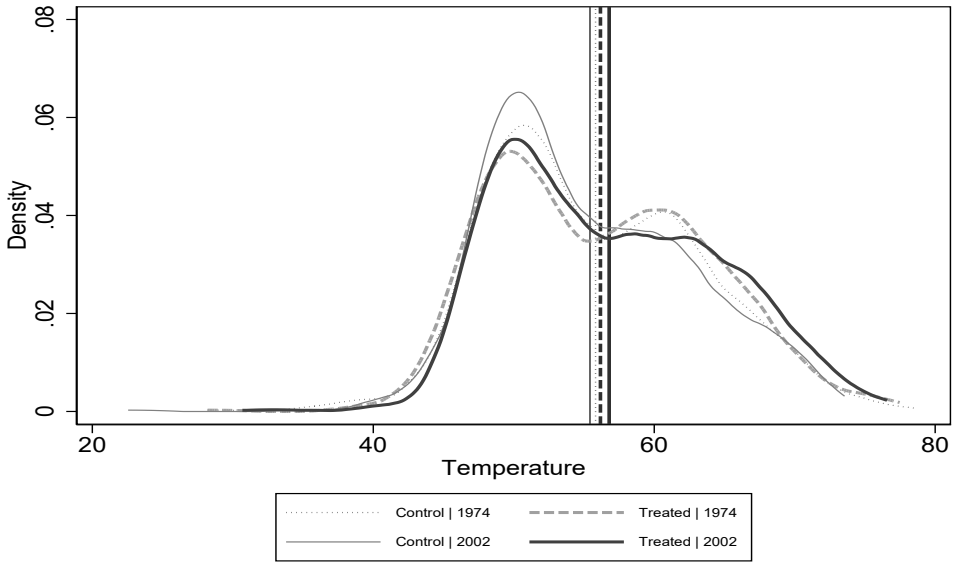
	(1) TEMP(Annual)	(2) TEMP(Winter)	(3) TEMP(Spring)	(4) TEMP(Summer)	(5) TEMP(Autumn)
DiD	0.421*	0.142	0.284	0.498**	0.211
	(0.24)	(0.36)	(0.22)	(0.22)	(0.33)
Treatment	0.039	0.107	0.051	-0.183	0.076
	(0.17)	(0.26)	(0.16)	(0.16)	(0.24)
Post	0.144	-0.777***	0.794***	0.283*	0.523**
	(0.17)	(0.26)	(0.16)	(0.16)	(0.24)
Latitude	-1.448***	-2.018***	-1.329***	-0.943***	-1.547***
	(0.01)	(0.02)	(0.01)	(0.01)	(0.02)
Wind Speed	-0.055***	-0.168***	0.004	0.063***	-0.093***
	(0.02)	(0.03)	(0.02)	(0.02)	(0.03)
N	2,767	2,763	2,765	2,766	2,766

(b) Flexible bounds

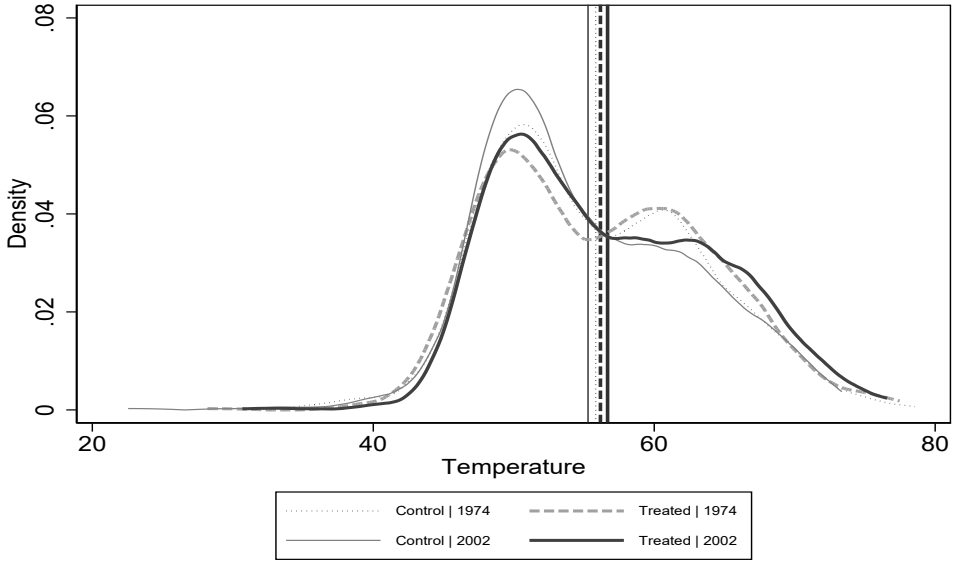
	(1) TEMP(Annual)	(2) TEMP(Winter)	(3) TEMP(Spring)	(4) TEMP(Summer)	(5) TEMP(Autumn)
DiD	0.294	-0.075	0.239	0.403*	0.010
	(0.25)	(0.37)	(0.23)	(0.23)	(0.33)
Treatment	0.054	0.128	0.052	-0.174	0.117
	(0.18)	(0.27)	(0.16)	(0.17)	(0.24)
Post	0.213	-0.647**	0.806***	0.356**	0.679***
	(0.18)	(0.27)	(0.17)	(0.17)	(0.24)
Latitude	-1.444***	-2.016***	-1.323***	-0.943***	-1.546***
	(0.01)	(0.02)	(0.01)	(0.01)	(0.02)
Wind Speed	-0.057***	-0.167***	0.008	0.071***	-0.097***
	(0.02)	(0.03)	(0.02)	(0.02)	(0.03)
N	2,735	2,732	2,731	2,731	2,733

Standard errors in parentheses
$*p < 0.1, **p < 0.05, ***p < 0.01.$

Effect of change in residential distribution and urban development on the Temperature of MSA's center. Difference in Difference method is based on Wave 1 (1972-1976) as pre-intervetion period and Wave 4 (2000-2004) as post-intervention. DiD coefficient reflects the causal effect and is statistically significant for Annual temperature, showing due to changes in sprawl patterns (horizontal development of cities) temperature of MSA's center is raised by 0.625 degree Farenheit. It also shows most of the effect is through the changes in temperature during Spring and Summer time. Validity of DiD results are based on strong assumptions of no unobserved heterogeneity bias and single treatment in long period of time (a) Control group is constructed to be the surrounding areas of the MSA center, fixed in time. (b) Robustness check of the results shown in Table 1.6a, using flexible bound scheme. Control group is constructed to be the surrounding area that varies depending on changes in residential distribution to reflect changes in city boundaries (flexible scheme).



(a) Fixed bounds



(b) Flexible bounds

Figure 1.8: Distributional effect of Treatment and Time in DiD

Source: Author's Calculation using Difference in Difference method

(a-b) Unconditional distributions of Treatment and Time dimensions are presented. . (a) Using the fixed bound scheme, it shows the average of the treatment group increases (moves to the right), while the average temperature of the control group decreases (moves to the left). (b) Like fixed bounds, the effect of treatment is exacerbated by time in flexible bound scheme.

the fixed and unobserved heterogeneities. As a result, this analysis is less likely to suffer from various sources of endogeneities, and the estimation results are more likely to reflect the causal effects of sprawl on the UHI effect. However, if the utilized instrument affects outcome directly, or in the presence of a weak instrument with low correlation with the endogenous variable (residential compactness), the results are not reliable and may suffer from grasping spurious relationships and reflecting overall data trends as causal relationships.

Table 1.7: F-stat of First stage of IV

	(1) Fixed Bound	(2) Flexible Bound	(3) Fixed Bound(FE)	(4) Flexible Bound(FE)
Coefficient	-0.072	-0.070	-0.038	-056
SE	0.002	0.002	0.004	0.007
F	715.223	596.550	124.903	117.947

First-stage F-statistics is conducted to show the validity of the second requirement for IV.

Table 1.8a shows the effect of residential compactness on the temperature of the urban area using the IV of planned highways, and MSA fixed effects using fixed bounds. The estimation includes all the observations for all four waves and 20 years (1972–2004). The IV estimates should be compared with both the fixed-effects estimates and DiD, as IV incorporates advantages of both methods. The IV effects are much larger than those calculated using fixed-effects estimation.

Since, compared with the fixed-effects method, IV is expected to reduce the reverse causality problem, an increase in effects suggests that two endogeneities (reverse causality and unobserved heterogeneity) are complementary and working in the same direction. That is, since after controlling for endogeneity the effects are increasing, the nature of reverse causality and unobserved heterogeneity is to reduce the observed effects and hence correlations do not show the severity of the effect. Table 1.8b shows the same result for the flexible

Table 1.8: IV effects on mean Temperature of urban area

(a) Fixed bounds

	(1) TEMP(Annual)	(2) TEMP(Winter)	(3) TEMP(Spring)	(4) TEMP(Summer)	(5) TEMP(Autumn)
RC	-30.996***	8.791	-41.833***	-48.758***	-37.801***
	(8.83)	(11.19)	(7.87)	(10.89)	(11.08)
Wind Speed	0.018*	0.014	0.019**	0.004	0.005
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Population	-0.730***	-0.240	-0.774***	-1.076***	-0.688**
	(0.26)	(0.32)	(0.22)	(0.31)	(0.32)
N	3,063	3,100	3,100	3,099	3,097

(b) Flexible bounds

	(1) TEMP(Annual)	(2) TEMP(Winter)	(3) TEMP(Spring)	(4) TEMP(Summer)	(5) TEMP(Autumn)
RC	-23.788***	10.873	-39.478***	-38.398***	-33.640***
	(6.35)	(8.18)	(5.75)	(7.69)	(8.13)
Wind Speed	0.009	0.007	0.006	0.003	-0.006
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Population	-0.617**	-0.286	-0.845***	-0.957***	-0.668**
	(0.25)	(0.30)	(0.21)	(0.29)	(0.30)
N	2,921	2,994	2,994	2,993	2,992

Standard errors in parentheses
$*p < 0.1, **p < 0.05, ***p < 0.01.$

IV estimation on the Heat Island Effects for Waves 1 through 4 (1972-2002). RC (residential compactness) coefficient shows the statistically significant causal relationship. Employed instrument is the planned portion of highways. As development of highways motivates individuals to relocate to the new suburbs and surrounding areas of the MSA center, scatteredness increases and residential compactness decreases. Hence, negative coefficient of RC is in accordance with the Urban Heat Island effect hypothesis which predicts, developing horizontal constructions increases the temperature of the central parts in cities. As residential compactness decreases, temperature of the MSA center increases. IV estimates of the causal relationships are significant for all seasons of year. However, for winter this relationship is positive and for autumn it is less significant than spring and summer. Thus, IV estimates reflect what was produced by DiD estimates and further, shows more extreme weather should be expected for the MSA centers during winter and summer, as cities develop horizontally. (a) Results using the fixed Bound Scheme. (b) Results using flexible bound scheme. These IV estimates of the causal relationships are significant for most of the seasons. They are closer to DiD effects comparing with the IV estimates with fixed bounds. Like fixed bound scheme, they show positive coefficient for winter but statistically insignificant. The effects are generally smaller than those produced using same method and fixed bound scheme (Table 1.8a).

Table 1.9: IV effect on mean Temperature of rural area (Control group)

(a) Fixed bounds

	(1) TEMP(Annual)	(2) TEMP(Winter)	(3) TEMP(Spring)	(4) TEMP(Summer)	(5) TEMP(Autumn)
RC	-7.183	21.802***	-14.305***	-14.401***	-11.093*
	(5.24)	(7.81)	(5.48)	(5.27)	(6.44)
Wind Speed	-0.007	-0.001	-0.010	-0.018**	-0.015
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Population	-0.091	0.445	0.215	-0.036	0.099
	(0.18)	(0.27)	(0.19)	(0.18)	(0.22)
N	3,672	3,670	3,671	3,671	3,667

(b) Flexible bounds

	(1) TEMP(Annual)	(2) TEMP(Winter)	(3) TEMP(Spring)	(4) TEMP(Summer)	(5) TEMP(Autumn)
RC	-7.510*	20.417***	-17.490***	-12.787***	-14.456***
	(4.13)	(6.10)	(4.11)	(4.12)	(5.18)
Wind Speed	-0.013*	-0.006	-0.006	-0.015**	-0.024***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Population	-0.058	0.363	-0.005	-0.020	0.018
	(0.18)	(0.26)	(0.18)	(0.18)	(0.22)
N	3,471	3,469	3,468	3,468	3,466

Standard errors in parentheses
$*p < 0.1, **p < 0.05, ***p < 0.01.$

IV estimation on the temperature of the outer city areas for Waves 1 through 4 (1972-2002). It is expected that sprawl has a less remarkable impact on the sounding area. RC (residential compactness) coefficients show a statistically significant causal relationship within the warmer seasons. However, the overall effect is insignificant. The employed instrument is the planned portion of highways. (a) Sensitivity analysis using fixed bound scheme. (b) Sensitivity analysis using fixed bound scheme.

bound scheme. The results in this scheme closely follow the results from the fixed scheme, meaning flexible bounds do not deviate much from fixed bounds. The result confirm that sprawl has a significant role in tempreture. The result confirms that sprawl has a significant role in the rise of the temperature of urban areas.

1.5.6 The Impact of Sprawl on UHI

In the previous section, we discussed the impact of sprawl on the temperature around MSA centers. However, sprawl may affect rural areas as well. Developing highways, even in places where they do not stimulate residential relocation, may increase the temperature by conducting traffic and industrial activities through surrounding MSA areas and, consequently, affect the temperature of not only the MSA center, but the surrounding areas as well. To examine this effect, I estimate IV coefficients for the temperature of the out-bound areas.

The results for the fixed scheme are shown in Table 1.9a and those for the flexible scheme are shown in Table 1.9b. These IV estimates indicate that the control groups (out-bound areas) are affected by the exogenous changes in residential compactness in both schemes. In other words, the instrument affects the outcome directly. However, comparing these estimates with the IV estimates of the in bounds, it is obvious that the IV estimates of the control group Tables 1.9a and 1.9b) are much smaller than those for treated group (Tables 1.8a and 1.8b). While an increase in residential compactness (opposite direction of sprawl) does have a negative effect on all seasons except for winter, The overall impact is only significant at 10 percent for the flexible scheme since the discretion of impact for winter has a positive sign. However, In general, the magnitude of the coefficient is smaller than those for urban areas, which means that sprawl indirectly affects the surrounding rural area but to a lesser degree. The estimates show that growth in constructed residential compactness has a negative and significant effect on temperature in all seasons. In other words, sprawl significantly contributes to the rise of the temperature.

Table 1.10: IV effects on UHI

(a) Fixed bounds

	(1) TEMP(Annual)	(2) TEMP(Winter)	(3) TEMP(Spring)	(4) TEMP(Summer)	(5) TEMP(Autumn)
RC	-30.933***	-22.906***	-36.559***	-36.645***	-27.219***
	(10.06)	(7.09)	(7.77)	(10.37)	(8.66)
Wind Speed	0.029***	0.015**	0.028***	0.013	0.014
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Population	-0.767**	-0.814***	-1.226***	-1.118***	-0.904***
	(0.30)	(0.20)	(0.22)	(0.30)	(0.25)
N	3,063	3,098	3,099	3,098	3,092

(b) Flexible bounds

	(1) TEMP(Annual)	(2) TEMP(Winter)	(3) TEMP(Spring)	(4) TEMP(Summer)	(5) TEMP(Autumn)
RC	-21.825***	-17.034***	-27.156***	-31.314***	-17.563***
	(7.16)	(5.18)	(5.18)	(7.44)	(6.28)
Wind Speed	0.023**	0.019***	0.013**	0.018*	0.017**
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Population	-0.683**	-0.726***	-0.972***	-0.984***	-0.766***
	(0.28)	(0.19)	(0.19)	(0.28)	(0.23)
N	2,921	2,993	2,992	2,991	2,987

Standard errors in parentheses
$*p < 0.1, **p < 0.05, ***p < 0.01.$

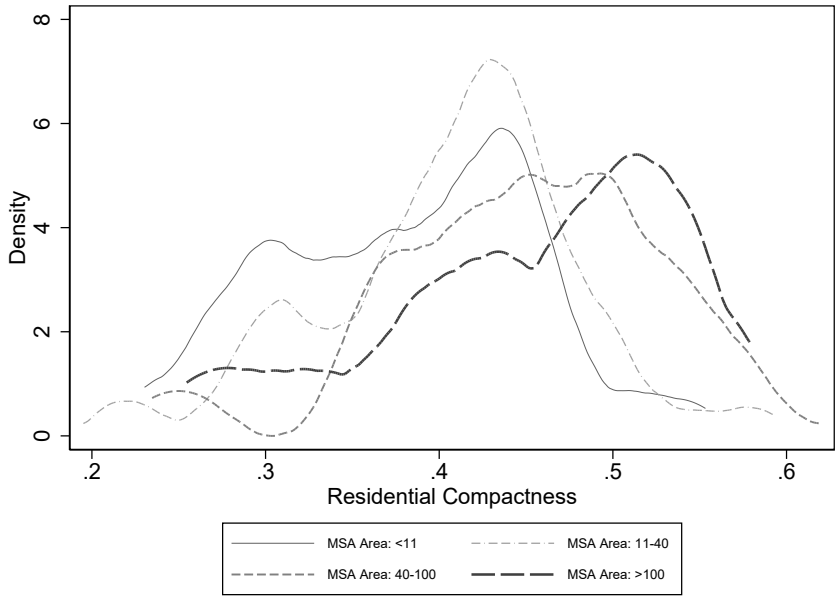
Causal effect of vertical development of the urban areas on difference in temperature of innercity and surrounding areas with MSA specific fixed-effect. This is to address UHI effect for Waves 1 through 4 (1972-2002). RC (residential compactness) coefficient shows the statistically significant causal relationship. However, compared with IV with fixed-effect results, reported in Tables 1.9a and 1.9b, marginal effects (RC coefficients) are smaller. Employed instrument is the planned portion of highways. As development of highways motivates individuals to relocate to the new suburbs and surrounding areas of the MSA center, scatteredness increases and residential compactness decreases. Hence, negative coefficient of RC is in accordance with the Urban Heat Island effect hypothesis which predicts, developing horizontal constructions increases the temperature of the central parts in cities. As residential compactness decreases, temperature of the MSA center increases. (a) Results using the fixed Bound Scheme. (b) Robustness check using flexible bound scheme.

Table 1.10a and 1.10b reflect the effect of residential compactness on the urban heat island effect by utilizing the difference between surrounding rural area and urban area as outcome and following IV estimation with fixed-effects. Flexible bound scheme leads to the smaller effects comparing to the fixed bound scheme. This can be due to the less tangible definition of outer city in flexible bound scheme as the geographical area associated with the control group moves in this scheme, and hence, coefficients are less substantial than those in the fixed scheme. It also magnifies the problem of interpreting the flexible scheme results that was discussed in Section 1.5.1. As an instance, the measure of residential compactness in Atlanta changes from .0195 in 1972 to .0065 in 2014, which shows $\Delta RC_{Atlanta} = -.013$. Using estimated coefficient of -30.933 for residential compactness in fixed bound scheme, I calculate the UHI effect that is caused by the increase in scatteredness to be, $E_{\Delta RC|1972-2014} \approx 0.43°F$ which explains almost half of rising in the temperature during this period. The same effect for flexible scheme is $0.28°F$.

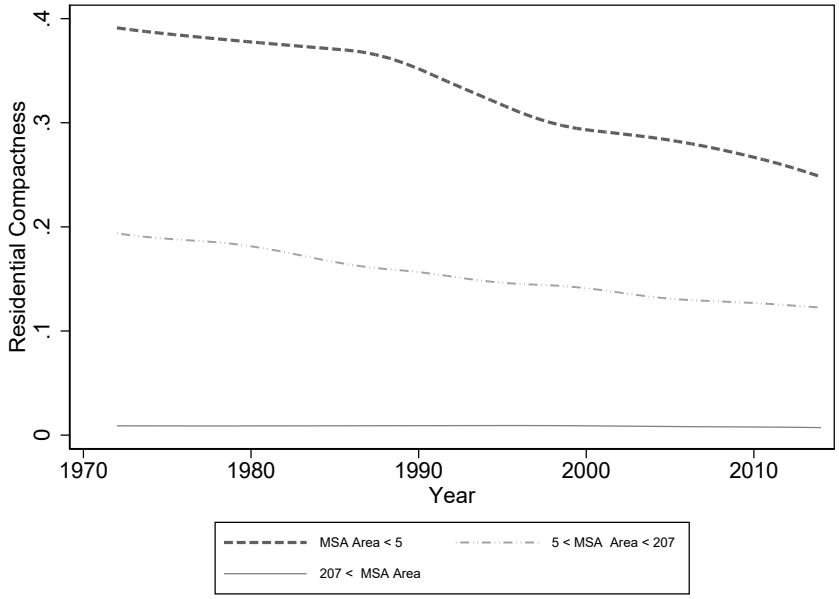
1.6 Robustness Check: Subsample Estimation

The selection of the estimation sample is based on the rationale that MSAs that are already developed and very small MSAs do not contribute to any of the results. **Figure** 1.9a shows the distribution of the residential compactness for each category of MSA size, divided into four categories, approximately representing quartiles of MSA size distribution. It suggests that relatively larger MSAs are more likely to be residentially compact, and hence it is less likely for them to be developed.

Most of large MSAs are located by the sea and are at the intersection of major trade routes, and hence the development of new highways is unlikely to affect them. On the other hand, for the smallest MSA group, their residential compactness measure is relatively smaller, suggesting that these MSAs are horizontally developed, and any development in them is vertical. However, based on the sample, these MSAs do not reach the vertical development level of the 50% middle MSAs by year 2014, which is the last year in the



(a) Distribution of Residential Compactness for each category of MSA size



(b) Changes in the Residential Compactness in time for different MSA sizes (Fixed bounds)

Figure 1.9: MSA size and Residential Compactness

Source: Author's Calculation using NWALT data and ArcGIS

(a) Distribution of the adjusted measure of Residential Compactness is depicted for each category of MSA area in 1950. While we might expect to observe negative relationship between area and Residential compactness, positive correlation is apparent. It suggests, development of large MSAs has been happened before 1950 and by the time, large MSAs had been representing highly dense area, as is reflected in Residential compactness. (b) Evolution of Residential Compactness in time for three categories of MSA sizes. Three categories which are depicted, are associated with 25th, 50th, and 75th percentiles. Smallest category shows the fastest increase in scatteredness (reduction in residential compactness) compared to the other two categories.

sample. These smaller MSAs are also located near the sea, and it seems that excluding either the MSAs near the sea or keeping the middle 50% leads to approximately the same result. Figure 1.9b shows changes in the adjusted[4] measure of residential compactness for each of the three categories of the smallest 25%, the middle 50%, and the upper 25% MSAs. Table 1.11 shows the result for subsample estimation.

Table 1.11: IV result of Effects on mean Annual Temperature for subsample of data

	(1) Inner city	(2) Outer city	(3) Difference
RC	-13.153***	-5.170	-12.183***
	(6.31)	(4.26)	(7.25)
N	3,063	3,672	3,063

Standard errors in parentheses
$*p < 0.05, **p < 0.01, ***p < 0.001.$

1.7 Conclusion

While the rise in the temperature of cities has been triggered by global warming and accelerated urbanization, not much is known about the causal effect of the different forms of urban development on the environmental features of cities. In other words, questions such as the environmental costs of vertical versus horizontal development of urban areas, or the effect of sprawl on the UHI effect, are open empirical questions without a clear answer. To provide clear and causal answers to these empirical questions, a range of issues needs to be addressed, from the lack of a precise definition for sprawl to the endogeneity and difficulties in defining the city horizon.

Accordingly, this study adopts the procedure developed by Burchfield et al. (2006) to construct a new measure of sprawl. A comprehensive measure of sprawl in this study allows for the incorporation of more MSAs and increases the statistical power of the tests. Also, utilizing the stations' information on the surface allows for detailed, daily aggregated surface data, such as daily temperature, the dew point, and the wind speed.

[4] Adjustment is to make a ratio from the count of the raster points each representing residential land and total area in MSA.

To address endogeneity, this study employs the IV method. It uses the planned portion of state highways as an instrument for sprawl. This variable was first developed by Baum-Snow (2007) as an instrument for interstate highways to study its impact on sprawl itself. Also, to better track the problem, I construct a flexible bound scheme that can be used in other studies where the entity under observation changes size through time, such as cities, natural currents, and resources such as woods.

The result suggests that sprawl can be responsible for substantial rise in the UHI effect. If we do not take into account the reverse causality, endogeneities can obscure the results severely.

Several studies have discussed the impact of UHI on heat-related outcomes, such as energy consumption and heath-related illnesses. However, further studies on the direct effects of urban configuration and sprawl on such outcomes are needed. While this study's focus is temperature, it makes way for future studies by constructing a clear measure of sprawl and combining it with an identification strategy that allows for multidimensional analysis of other consequences of urban sprawl such as infrastructure costs, inequality, segregation, and public health.

Chapter II

The Role of Labor Unions in Response to COVID-19 Pandemic

2.1 Introduction

This study investigates the role that labor unions, one of the largest institutions in the United States, play in the spread of COVID-19.[5] Person-to-person transmission in the workplace is thought to play a crucial role in the spread of the virus. While a complete shutdown of businesses is neither possible nor optimal for an extended period of time, in the absence of a vaccine marginal alterations in work schedules and appropriate workplace safety measures are vital to the success of efforts to control the spread of the virus. Unions play a role in shaping and regulating employer–employee relations. The direction and significance of the effect of unions on the spread of the virus, however, remains an empirical question. This paper analyzes these effects by utilizing state-level data in the United States and by employing a dynamic nonlinear probability model and method of moments (MM) estimation.

Freeman and Medoff (1984) distinguish between two dimensions of unions: the monopoly face and the collective voice/institutional response face. Through their monopoly power, unions can affect the spread of COVID-19 by imposing employees' preferences through the alteration of layoff patterns, compensation benefits, and work conditions. The collective voice face, on the other hand, refers to the institutional impact of unions. Unions increase communication between employees and employers and facilitate the preferences of each to be revealed to the other. Preference revelation itself affects the level of economic activity and safety regulations at the workplace. Also, since the safety provisions and working schedules display characteristics of a public good, an adequate level may not be reached in an indi-

[5] Around 11% of employees in the United States are represented by unions.

vidual agreement between an employee and employer [Flanagan (1983)]. Noting the public good's aspects of the efforts to stop the spread of the virus, the collective voice face of unions not only may impact the spread of COVID-19 through higher nonwage benefits for workers which were supposedly shaped before the pandemic, but it can also play a role in adjusting working conditions during the pandemic by both reflecting and enforcing the optimum level of the costly efforts that satisfy the welfare function of union members.[6]

The present study utilizes lagged labor market data (to avoid simultaneity bias) and provides new evidence of positive externalities for union employees compared with nonunion employees. It suggests that increasing union size by 1,000 in the United States would lead to approximately 110 fewer cases of COVID-19 11 months after the onset of the virus.[7] Since we expect union members to have a more powerful voice than nonunion employees, if there are inconsistencies between employees' weights over consumption and health and those of the government or firms, a comparison between unionized and nonunionized employees reveals those inconsistencies. By utilizing this comparison, we also find evidence of better optimization by unions rather than state or federal governments, emphasizing the role of better information flows in smaller organizations with a powerful voice derived from collective bargaining power.

Uncovering the effect of unionization on the spread of COVID-19 contributes to the strand of economic literature that studies the externalities of labor unions from the insider-outsider point of view; this suggests that unions can help nonunion employees and the unemployed by slowing the spread of the virus. Understanding the role organizations such as unions play in efforts to prevent the spread of contagious viruses and helps to implement better policies in response to possible future pandemics.

[6] The International Union, United Automobile, Aerospace and Agricultural Implement Workers of America (UAW) claims to monitor and assess the situation while engaging with workers to help address some of the workplace issues that the COVID-19 crisis has presented. The statement is available at: uaw.org/coronavirus

[7] The results can be translated to a 10% increase in unionization from 11% to 12.1% of the U.S. employed wage and salary workers would lead to approximately 500,000 fewer cases of COVID-19 315 days after the onset of the virus. The total number of cases at day 315 (end of November) is 13,534,334, which suggests approximately a 3.7% decrease in total cases by the end of November 2020.

The remainder of the paper is organized as follows. After describing the mechanisms through which unions may affect the spread of the virus in section 2.2, in section 2.3 we discuss the estimation method and the model employed. Section 2.3.2 discusses the identification strategy, and section 2.4 describes the data. Section 2.5 provides the results, and section 2.7 concludes the paper.

2.2 Mechanisms

This section discusses the validity of the study's question by analyzing how unions can affect the spread of COVID-19. We investigate the possible effects of unions on the decision-making process of individuals and firms facing the pandemic. The effects may vary depending on the different union's faces: monopoly face and institutional face. We further discuss how unions can play a similar role to the government when protective measures in response to COVID-19 reflect the public good's characteristics. Lastly, we address different externalities the union's action may have.

Employees and firms (as economic agents) play an essential role in determining the transmission rate of COVID-19 by making multiple work-related decisions. These decisions can be broadly categorized into decisions about the level of safety measures and economic activity. We expect that by reducing economic activity and increasing safety measures, the spread of the virus will slow down. However, both measures are costly to agents. Also, at least in the short term, there is a trade-off between the level of economic activity and occupational safety in the event of a pandemic. Hence, the level of restrictive measures should be optimized, considering weights of consumption (employment) and health in employees' and firms' utility and profit function, respectively. In the absence of unions, federal and state governments are the only agents affecting economic agents' decision-making processes. They do this by setting social distancing restrictions and advising businesses on how to provide a safe workplace. Unions as a form of collective bargaining in employer–employee relations (as opposed to individual bargaining) can affect the rules and effectiveness of social distancing

restrictions. The effect may vary depending on the monopoly power of unions and their efficiency in revealing employers' and employees' preferences to one another.

Depending on the monopoly power they possess, unions affect the bargaining power of their members (enforcement mechanism) and reduce information transmission costs (preference revelation) between employers and employees [Boxall and Purcell (2011)]. They provide their members with better contracts thereby guaranteeing more job security and better wages. Unions increase the probability of their members receiving employer-provided health insurance [Buchmueller et al. (2002)]. Ninety-one percent of unionized workers can take paid sick leave compared with 73% of nonunionized workers [Pizzella and Beach (2019)]. Unionized workers are slightly more likely to have paid leave, and their paid leave benefits are significantly higher in dollar value [Mishel et al. (2012)]. Also, they are more likely to receive employer-provided pensions and health insurance with a far larger impact on the magnitude of benefit [Pierce (1999), Budd and Na (2000)]. Longitudinal studies also show that a decline in union density explains approximately one-fourth of the decline in aggregate health insurance and pension coverage [Bloom and Freeman (1992); Strombom et al. (2002)). Since efforts to control the spread of the virus are costly for both employers and employees, these factors may help union members to better protect themselves from contagious diseases by shifting protection costs to firms (by utilizing medical services, decreasing work hours, and increasing work safety measures while remaining employed).

With regard to the preference revelation mechanism, unions can affect protection levels. Since information about an individual's health is private, a lack of information prevents the employer from reaching an optimal decision in terms of the level of job safety necessary for operation. In such cases, an employer may ask its employees to attend an unsafe workplace, or conversely, implement unnecessary regulations at work. Unions, in this case, increase transparency by revealing employees' health levels to the firm, and also by informing employees about their rights, their contracts, and the perspective of the whole industry. In the absence of a union this information would be costly for employees to acquire. Thus, higher

unionization levels can be helpful in controlling the contagious disease if unions inform employees about their rights, and this could lead to better workspace safety. In contrast, if the union's evaluation of the future of the industry is pessimistic, it can help union members and employers to cooperate by attending a less safe workplace and keeping production costs low, which leads to maintaining union members' jobs and the firm's profit currents. As a result, unions can positively or negatively affect the spread of the virus from the perspective of transparency and trust among employees and employers. Also, apart from the efficiency of unions in revealing the employees' preferences, unions, due to their collective voice nature, can determine which employees' preference being revealed to the employer. Freeman (1981) suggests that unions reflect median voter preferences. While the median worker has a higher demand for nonwage benefits, nonunionized firms provide compensation in the form of bonuses to the marginal worker.

Considering that in the case of a contagious disease such as COVID-19, the benefits of slowing down the spread of the virus are not limited to employees and union members, it is plausible to consider health and better protective measures as a public good. Without an enforcement mechanism, public goods like preventive measures would be undersupplied. Higher levels of such public goods could yield higher social and individual welfare (which we assume is guiding the government's interventions). Since the safety of the work environment is a public good, it is unlikely that it will be provided sufficiently by firms without external pressure. As a result, some governmental intervention in the competitive solution is justifiable. Weil (1999) shows these type of interventions are more of a union supplement than a substitute, and unions play a substantial role in the enforcement of the Occupational, Safety, and Health Act (OSHA) in the manufacturing sector [Weil (1991)].

While both governments and unions optimize safety measure levels and economic activities by choosing health and employment levels, they may differ in at least two aspects: the information set they possess and the objective function they target. Ideally, the government optimizes the social welfare function and incorporates broad information about individual

preferences. On the other hand, unions are expected to optimize the same aspects for their members only, using a more narrowly defined information set, one limited to their members. These differences in the scope of the objective functions, and the quality and volume of information, raise the questions of how and to what extent unions are contributing to contain the spread of the virus. Owing to the rapidly growing and unprecedented nature of the COVID-19 pandemic, the government's information about the weight individuals assign to their health is limited, and thus the government's objective function may not reflect the social welfare function. Unions have more detailed information with regard to their members' preferences as a result of having closer contact with them. This means they are more effective in both collecting information about preferences and imposing restrictions. However, unions are not necessarily targeting the same level of restrictions as governments. The characteristics of union members can affect the union's objective function. Collective bargaining may increase the probability of infection if the collective decision of the union is in favor of working more (e.g., in cases where individuals believe the disease is less dangerous for them than it is for the average individual in the community).

As discussed previously, due to the public good's characteristic of health in case of a contagious disease, the effect of changes in workplace safety measures and work schedules of union members can also be transmitted to other people and hence have positive or negative externalities. Externalities derived by union's actions regarding COVID-19 can happen in at least three ways. First, unions alter the probability of union members being exposed to COVID-19, which itself affects the spread of the virus. Second, changes in work schedules of the union members can affect the work schedules of other employees through general equilibrium response. As an example, a change in prices due to the union members working less may encourage other employees to work more. Combined with the level of safety measures of nonunionized employees, this affects the probability of nonunionized being infected by the virus. Lastly, the effect of unions on working conditions is not limited to union members or unionized workplaces. Unions set standards for work conditions that could be adopted by

the labor market in general [Western and Rosenfeld (2011)]. As a related example, higher union density rates are associated with lower levels of economic inequality [Alderson and Nielsen (2002); Alderson et al. (2005); Atkinson (2003); Western and Rosenfeld (2011); Neal (2013)]. Also, the collective bargaining power of unions is not limited to the firm itself. Bargaining takes place at upper levels such as industry, state, and national levels. Hence, unions contribute to pushing through legislation on social programs that impact society in general, such as national social security, unemployment compensation, and minimum wage laws [Asher (2001); Galenson (1986)].[8]

In summary, unions may play a role in containing or spreading a contagious disease, depending on their monopoly power and preference revelation efficiency. Hence, determining the aggregate effect of unions on the spread of a contagious disease is an empirical question that we discuss in what follows.

2.3 Model and Estimation

2.3.1 Introduction to Model

To evaluate the effect of unions on the spread of COVID-19, we require a predictive dynamic model that also incorporates the treatments controlling for union membership and involvement in economic activities. Our modeling approach is based on estimating an individual's probability of infection in a dynamic framework, where probabilities are a function of treatments, among other variables. It is achieved by incorporating a logistic model for daily individual infections and aggregating the results to reach the state level's total number of infections.

The assumption behind aggregating the individual probabilities is independence between the individual probabilities. To address the dependencies between individual observations

[8] In the case of COVID-19, The American Federation of Labor-Congress of Industrial Organizations (AFL-CIO) has been actively negotiating OSHA, Congress, federal agencies, and state and local governments during the pandemic. AFL-CIO's statement is available at: aflcio.org

(probability of infection for one person changes as other people become infected), we integrate elements of the compartment modeling approach into our model. Compartment modeling, which was formulated by Kermack and McKendrick (1927), is well studied in the epidemiology literature and concerns the prediction of the spread of contagious diseases. We borrow the elements of a simple compartment model in which there are two compartments between which individuals are allowed to move. Individuals are either susceptible or infected at every point in time. Using the state as the relevant geographic unit, the probability of being infected (conditional on not being infected at the time) is a function of the total number of infected in the state in the previous week. Next, we multiply the individual probability of being infected in each state by the total number of susceptible individuals in the state to reach the total number of new cases. In other words, by controlling for the spread of the virus, provided with the elements of compartment modeling, we can use a logistic probability where we control for the dependence between individual probabilities. To improve predictive power, we further control for various state-level characteristics and trends. Lastly, we augment the predictive model with another set of variables that address the identification of the marginal effect of interest, including the state-level characteristics of union members versus all employees and variables to isolate the treatment variable itself. Our probability model can be written as:

$$y_{jd} = S_{jd} \times Pr(\zeta_{jd}) + \epsilon_j \tag{2.11}$$

where y_{jd} is the number of new cases, S_{jd} is the number of susceptible individuals. and $Pr(\zeta_{jd})$ is an individual probability of being infected as function of characteristics (ζ_{jd}) in state j and day d. The ideal probabilistic model should use individual-level observations to address the characteristics. In a situation where the outcome is available at the state level, using individual-level characteristics imposes a high level of computational expenses. To avoid these computational complexities, we use state-level averages.[9] In other words, ζ_{jd} is a linear combination of the state j and day d average socioeconomic characteristics

[9] This assumes that the state-day average reflects the individual-day characteristics and ignores the bias

$\vec{X}_{jd}$, the set of time invariant state-level observable variables $\vec{\Gamma}_j$,[10] time trends $\vec{D}_d$ which intends to reflect the seasonal COVID-19 trends, $\vec{G}_{jd}$ that includes lagged total number of cases and neighboring effects, the set of treatment variables $\vec{T}_{jd}$, occupation shares $\vec{\Sigma}_{jd}$, and lastly, social distancing restrictions which are set by the government R_{jd}. The linear index is defined as:

$$\zeta_{jd} = \vec{\bar{X}}_{jd}.\vec{\beta}_X + \vec{\Gamma}_j.\vec{\beta}_\Gamma + \vec{D}_d.\vec{\beta}_D + \vec{G}_{jd}.\vec{\beta}_G + \vec{T}_{jd}.\vec{\beta}_T + \vec{\Sigma}_{jd}.\vec{\beta}_\Sigma + R_{jd}.\beta_R \tag{2.12}$$

For each state-day, we use the Current Population Survey (CPS) monthly data to calculate the average socioeconomic characteristics ($\vec{X}$), treatments ($\vec{T}$), and occupation share variables ($\vec{\Sigma}$). Due to the CPS limitations, all of them are state-specific, monthly variables. That is, their value may only update once per month. Our decision to use "day" as the unit of time combined with employing monthly variations leads to regression dilution. In other words, since monthly variables do not vary during the month, it is as if we have a measurement error for monthly variables, which leads to attenuation bias.[11] For socioeconomic characteristics, we have:

$$\vec{\bar{X}}_{jd} = \frac{\sum_{i \in j} \vec{x}_{id}.w_{id}}{\sum_{i \in j} w_{id}} \quad \forall j \tag{2.13}$$

where $\vec{x}_{id}$ is a vector of socioeconomic characteristics for individual i in state j and day d,[12]

due to Jensen's inequality that is in play in the presence of a nonlinear probability function. In other word:

$$Pr(E[\zeta_{jdi}]) \neq E[Pr(\zeta_{jd})]$$

where index jdi refers to individual i in state j and day d.

[10] We also use the second power of variables in $\vec{\Gamma}$ with enough variations.

[11] While we are aware of this issue, we think avoiding this by aggregating over the outcome (COVID-19 cases) is even more problematic. For example, the month as the time period (instead of the day) is too long that it prevents proper incorporation of the restriction rules and inter-state transmission. The latter one implies a violation of observational independence. Also, simple aggregation of the COVID-19 cases ignores the fact that the effect of a new case varies depending it happened early in a month or not. This inflates the estimate of growth rate (lagged outcome) in compartment modeling and leads to overestimation bias at the early stages of the pandemic by putting more weight on later days in each month.

[12] Including age, education, family income, living in metro area, married, and male. Hence we have:

and w_{id} is sampling weight. We further augment the socioeconomic dimensions by a set of variables which are available in state-level and set at their pre-COVID levels, $\vec{\Gamma}_j$ such that:

$$\begin{aligned}\vec{\Gamma}_j =\{&\text{Local self-dependence}_j, \text{Residential compactness}_j,\\ &\%\text{working from home}_j, \text{Political preference}_j,\\ &\text{Health}_j\}\end{aligned}$$

$\vec{D}_d$ includes seasonal time trends.[13] $\vec{G}_{jd}$ consists of the total number of cases in previous week MA(7) and a neighboring effect measure λ_{jd} that reflects the spread of the virus in surrounding states:

$$\lambda_{jd} = \sum_{k \neq j} \frac{y_{j(d-7)}}{(L_{kj})^4} \tag{2.14}$$

where y_{jd} is the number of new cases in region j on day d, and L_{kj} is the distance between regions j and k. The neighboring effect λ controls for the contamination between state-level observations. As a result, an increase in the number of cases in a neighboring region affects region j, where the same increase in a region distant from region j has a smaller impact.

$\vec{T}_{jd}$ represents the treatment variables. Since we are interested in the effect of a change in both extensive and intensive margins of the organized labor force, both unionization level and hours of work in the previous week for unionized employees are included. To compare the results with those of the all the employed labor force, we also include employment levels and hours of work in the past week for all employees. We also control for the share of organized employees and all employees in each of the six occupation categories ($\vec{\Sigma}$) [14] to reduce the risk of contaminating the marginal effects with the possible concentration of unionized employees

$$\begin{aligned}\vec{\vec{X}}_{jd} =\{&\text{Age}_{jd}, \text{Education}_{jd}, \text{Family Income}_{jd},\\ &\%\text{Metro}_{jd}, \%\text{Married}_{jd}, \%\text{Male}_{jd}\}\end{aligned}$$

[13] Day in each season and their second power.

[14] Including Management, Professional, Nursing, Service, Sales & office, and construction. We exclude a category related to the "production, transportation, and material moving" occupations to avoid perfect multicollinearity

in the occupations with lower (higher) exposure to the COVID-19 virus. The construction of the treatment variables is similar to the socio-economic variables as explained in Equation 3. Hence, the treatment vector $\vec{T}_{jd}$ is defined as:

$$\vec{T}_{jdc} = \{\%\text{Employed}_{jdc} \quad , \quad \text{hours}_{jdc}\}$$

where index c refers to the targeted treatment. Depending on the model specification (further discussed in Section 2.5), c may include union members, those who are covered by unions but not a member, and all employed labor force. While there are multiple alternatives for incorporating social distancing restrictions into the model, we find the simple aggregation of the number of restrictions reliable:

$$R_{jd} = \sum_{\rho \in \varrho} r_{\rho,jd} \qquad , \quad \varrho = \{1, 2, 3, 4, 5\} \tag{2.15}$$

where $r_{\rho,jd}$ is restriction ρ in day d and state j. The set ϱ contains mass gathering restrictions, initial business closure, educational facilities closed, non-essential services closed, and stay at home order.

We use a MM estimator with a logistic function as the link function to estimate the set of coefficients $(\vec{\beta}_X, \vec{\beta}_\Gamma, \vec{\beta}_D, \vec{\beta}_G, \vec{\beta}_T, \vec{\beta}_\Sigma, \beta_R)$. Residual u_{jd} can be define as:

$$u_{jd} = \; y_{jd} - \frac{S_{jd}}{1 + exp\left(-\zeta_{jd}\left(\vec{\beta}_X, \vec{\beta}_\Gamma, \vec{\beta}_D, \vec{\beta}_G, \vec{\beta}_T, \vec{\beta}_\Sigma, \beta_R | \bar{X}_{jd}, \Gamma_j, D_d, G_{jd}, T_{jd}, \Sigma_{jd}, R_{jd}\right)\right)} \tag{2.16}$$

The predicted number of new cases can be decomposed into the probability of becoming infected by the virus for representing individuals, multiplied by the number of people prone to the virus (susceptible) in state j and on the day d. Then, an unweighted MM estimator

minimizes the following statement with respect to the coefficients $\vec{\beta}$:

$$\widehat{\vec{\beta}} = \underset{\vec{\beta}}{\operatorname{argmin}} \left(\frac{1}{N} \sum_{j \in J} \sum_{d=1}^{D} \vec{Z}_{jd}.u_{jd}(\vec{\beta}) \right)' \left(\frac{1}{N} \sum_{j \in J} \sum_{d=1}^{D} \vec{Z}_{jd}.u_{jd}(\vec{\beta}) \right) \tag{2.17}$$

where $\vec{Z}$ is the vector of all explanatory variables and N is the total number of observations.

Our modeling strategy has multiple advantages when compared to a dynamic linear probability approach. While the linear probability model is prone to the out-of-range predictions (negative number of cases or more than total number of susceptible population in each state), the prediction based on the non-linear probability lies in the allowed range by design. Also, due to the non-linearity provided by logistic function, marginal effects are local. In other words, the non-linear probability model weights the different values of treatments differently. Whereas in the linear probability model, this can be achieved only by adopting complex specifications at the risk of misspecification. That is, our approach is less prone to misspecification bias.

MM estimators are comparable to the Maximum Likelihood (ML) estimators with respect to the estimators' efficiency. Using non-linear residuals allows us to keep observations with $y_{jd} = 0$ (no new case in a particular state-day). This is in contrast with OLS estimators of the odds ratio transformation where the observations with odd equal zero have to be dropped.[15] See section 2.6 for more details.

2.3.2 Identification

There are two main identification concerns in estimating the effect of a change in unionization level on the spread of the COVID-19 virus. Firstly, since union membership is a choice, if the factors which affect the decision to become a union member also affect individuals' behavior in following the social distancing guidelines, exogenous change of unionization level might not reflect the correct effects (selection bias). In other words, the treatment is

[15] After analyzing the results using OLS estimation we conclude that assigning a small value (10^{-4} to 10^{-5}) to y_{jd} where there is no new case, the OLS results converge to MM.

not random and hence, we cannot isolate the effect of unionization. Secondly, if the unionization level is endogenously determined by the spread of the virus, an exogenous change in the unionization level to calculate the marginal effects no longer provides a correct counterfactual analysis (simultaneity bias).

The first identification problem concerns the bias due to the non-randomness in union membership. Union membership can be thought of as a constrained choice made by employees. Duncan and Leigh (1985) model union membership choice as a function of employee's characteristics and benefits of union membership. We showed in Section 2.3.1 that we control for those characteristics that can affect union membership and behavior of individuals in response to the spread of the virus. While we control for the possible sources of confounders, there is no guarantee that the utilized set of covariates is sufficient for all the relevant state-level determinants of the spread of epidemics that may be correlated with union membership. To address this concern, we analyze the effect of an increase in the total number of employees who have access to the union benefits but are not member of union. This group is called *covered nonmembers* as opposed to the union members. Since the covered employees have not made the choice to join labor union, it is plausible to assume that the distribution of treatment (union coverage) is independent of the employees' characteristics that potentially determines the response of individuals to COVID-19. Hence, by comparing the results of an increase in union membership with the same absolute increase in the number of covered nonmember employees we can identify if the results are derived from different behavior of employees (if the results are different) or the union coverage itself (if the results are the same). Our comparison shows the marginal effects induced by both groups of employees are same in direction while different from those marginal effects from all employees. This suggests that the set of controls we utilize to differentiate between employed and union workers is sufficient to a degree that marginal effects are orthogonal to the individuals' characteristics that affect union membership decision. We explore this comparison in more details in Section 2.5.

For the simultaneity bias problem, we utilize lagging explanatory variables following Green et al. (2005) and Spilimbergo (2009) among others. In particular we use 2019 data for the treatment variables ($\vec{T}$), individual characteristics ($\vec{X}$), and occupational shares ($\vec{\Sigma}$). Simultaneity bias can be less serious for the organized labor than non-organized labor since the unionized employees are less likely to being affected by exogenous shocks such as COVID-19. This can be explained by the rigidity in union membership and contracts that increase employers' layoff costs. Moreover, since union members usually have multi-year contracts, union wages are less flexible than is otherwise determined by the market [Kaufman (2004)]. Rones (1981) claims that these multi-year agreements are restrictive to the point that they encourage firms to use early retirement as a tool to manage workforce flow in times of recession.

As explained previously, we also control for hours of work in the previous week to further explore the channels through which unions contribute to the spread of the virus. This isolates the effect of union size. As a result, the marginal effect calculated for changes in union size better reflects the intrinsic role of unions.[16] Union members can use their collective voice to alter their work hours in response to the spread of the virus. If there is endogeneity between intensive margin and spread of the virus, calculated marginal effects for union size (extensive margin) may be biased.

2.4 Data

We utilize daily generated data on the number of cases and deaths due to COVID-19, collected by The New York Times (2020a). These include the number of new and total cases in each state-day. Figure 2.10 shows the geographical distribution of COVID-19 virus in the middle of each month from January to November. Beginning in January from west-coast, epidemic center has shifted to the new-england states in March and gradually has moved to the southern states.

[16] For example, the safety measures utilized in union versus nonunion establishments.

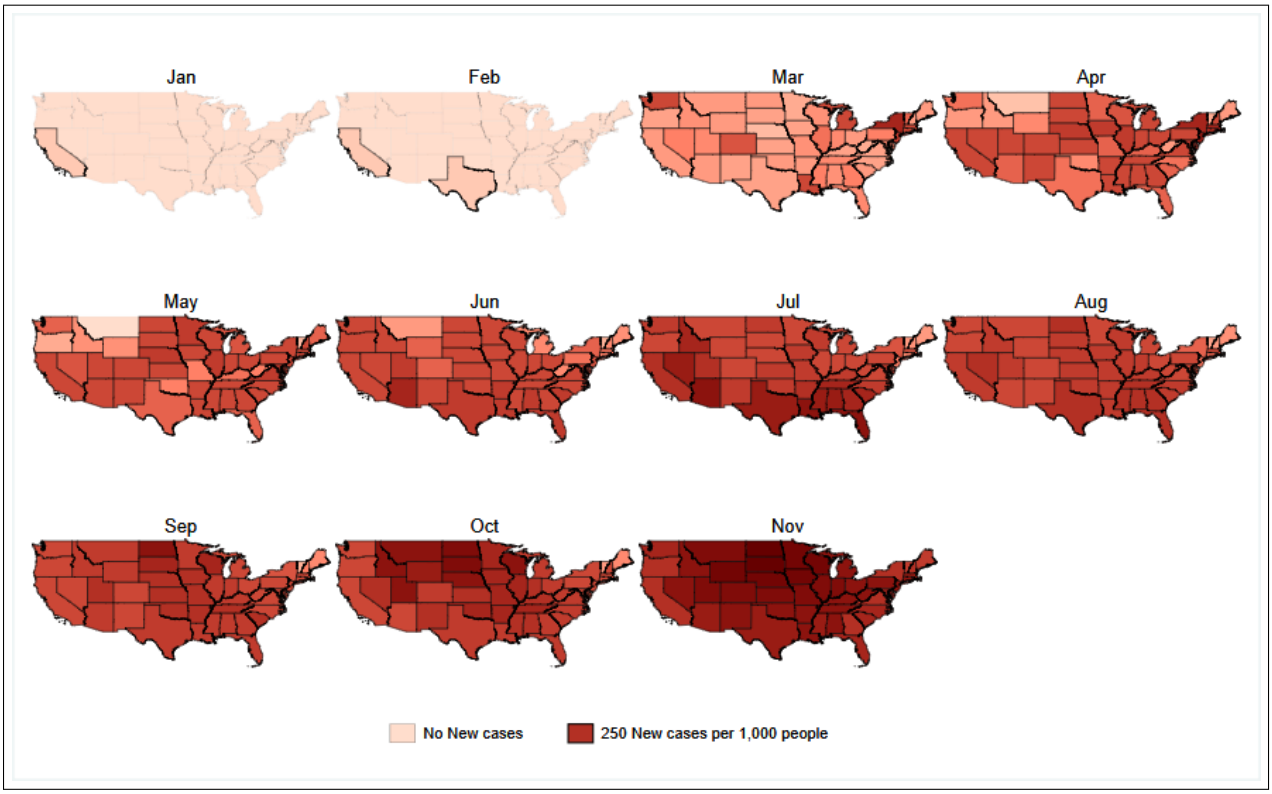


Figure 2.10: Number of new COVID-19 cases (observed and simulated)

The observed and simulated number of new cases of COVID-19 are shown in the graph for the whole United States and different census regions, as suggested by the U.S. Census Bureau. For each category, we provide simulation under the current level of unionization and simulation under the counterfactual level of unionization, where we increase the unionization level by 10%, keeping the employment level constant (Model 5 in Table B.1).

Socioeconomic ($\vec{X}$) and work-related status ($\vec{T}$ and $\vec{\Sigma}$) data are based on lagged CPS data to avoid simultaneity bias. We use monthly CPS data from January to November 2019 to extract the monthly state-level averages of the individuals' characteristics, family structure, and income. Work-related status includes employment status, hours of work in previous week, and share of each of the six occupation categories for union members, non-members who have union coverage, and total wage and salary employees reported by CPS. We follow Hirsch and Macpherson (2003) in calculating union status and use BLS weights, which also are used by the U.S. Bureau of Labor Statistics.[17] Table 2.12 reports the related summary

[17] Our results are robust to the choice of weight we use. Using household weights only slightly affects the results and neither changes the statistical nor economic significance of any of the main results.

of statistics for employed, unionized, and covered nonmember workers and compares 2019 with 2020 data.

Table 2.12: Summary of statistics.
Socio-economic ($\vec{X}$) and teartment ($\vec{T}$) variables

			Members	Covered	Employed				Members	Covered	Employed
$\vec{X}$	Age	2019	45.00	43.80	43.40	$\vec{\Sigma}$	Management	2019	0.07	0.12	0.17
			(3.28)	(8.68)	(1.24)				(0.06)	(0.19)	(0.03)
		2020	45.07	44.54	44.00			2020	0.07	0.14	0.18
			(3.40)	(8.25)	(1.37)				(0.07)	(0.21)	(0.03)
	Education	2019	2.19	2.42	43.80		Professional	2019	0.30	0.31	0.17
			(0.18)	(0.71)	(8.68)				(0.13)	(0.30)	(0.04)
		2020	2.41	2.56	44.54			2020	0.30	0.32	0.18
			(0.35)	(0.71)	(8.25)				(0.13)	(0.31)	(0.04)
	Income	2019	97.68	91.98	92.31		Nursing	2019	0.07	0.10	0.09
			(17.47)	(33.51)	(12.40)				(0.07)	(0.16)	(0.02)
		2020	103.09	97.23	97.78			2020	0.08	0.11	0.10
			(18.40)	(36.99)	(13.19)				(0.08)	(0.18)	(0.02)
	Living in Metro	2019	22.61	18.02	23.35		Service	2019	0.13	0.12	0.15
			(19.89)	(26.86)	(17.45)				(0.09)	(0.19)	(0.02)
		2020	22.13	19.81	23.54			2020	0.11	0.11	0.12
			(20.02)	(27.99)	(17.62)				(0.09)	(0.18)	(0.02)
	Married	2019	58.92	50.07	52.62		Sales & Office	2019	0.13	0.15	0.21
			(13.96)	(34.10)	(4.36)				(0.09)	(0.21)	(0.02)
		2020	60.62	48.61	53.86			2020	0.12	0.13	0.20
			(14.54)	(35.91)	(4.03)				(0.10)	(0.21)	(0.02)
	Male	2019	56.22	44.00	52.68		Construction	2019	0.14	0.08	0.09
			(13.05)	(33.22)	(1.52)				(0.09)	(0.15)	(0.02)
		2020	55.15	41.46	52.84			2020	0.13	0.08	0.09
			(14.19)	(35.40)	(1.57)				(0.09)	(0.15)	(0.02)
	Public	2019			9.30	$\vec{T}$	% Total	2019	12.19	1.60	61.54
					(2.31)				(6.26)	(1.07)	(4.16)
		2020			9.44			2020	11.82	1.51	57.99
					(2.54)				(6.22)	(1.15)	(4.44)
							Hours	2019	41.48	39.53	38.95
									(3.25)	(7.14)	(0.84)
								2020	40.31	39.28	38.19
									(3.28)	(8.27)	(0.92)

Standard errors are in parentheses.

Table 2.13 shows the summary of statistics for the variables which are not directly derived from CPS and are fixed in time at their pre-COVID levels ($\vec{\Gamma}$). These variables include health, political preference, working from home, and urban structure. To include state-level variations in health we use self-reported health from the Centers for Disease Control and Prevention (CDC)'s Behavioral Risk Factor Surveillance System (BRFSS) 2018 data. To incorporate a measure of the political preference we utilize share of democrats from the MIT

Election Data and Science Lab (2017).

Local self-dependence is based on the residential compactness and commercial accessibility measures which are provided in Rahimzadeh (2020). Representing the density of urban structure and commercial accessibility in each state, these two measures are constructed using Land use data and geographic information system (GIS). Commercial accessibility is the average percent of the commercial area around each residential cell and captures the employment accessibility. For constructing the local self-dependence, we use the residential compactness-defined as the average percent of residential land within a neighboring circle-and divide it by the commercial accessibility. Hence, local self dependence is larger if neighborhood is more compact or has a lower access to commercial centers and measures the degree of neighborhood clusteredness in each state. Associated with urban structure, is a working from home variable from the American Community Survey (2018). Lastly, we use the state-level restrictions repository from Institute for Health Metrics and Evaluation (2020). Table B.4 reports the detailed level of restrictions in each state.

Table 2.13: Summary of statistics.
Fixed in time variables ($\vec{\Gamma}$)

Residential Compactness	Local self dependence	Working from Home	Political preference (Democrat)	Health (More is poorer health)
0.40	6.24	4.30	0.45	2.55
(0.06)	(1.18)	(1.05)	(0.12)	(0.11)

Standard errors are in parentheses.

2.5 Results

In this section we analyze the effect of extensive and intensive margins of labor supply for union members, covered nonmembers and all employees, on the total number of confirmed COVID-19 cases. Marginal effects are based on the simulation of the daily spread of the virus. In what follows we discuss the calculation of the marginal effects and provide the results.

Lastly, we address the endogeneity concerns due to the selection in union membership and simultaneity bias.

In a dynamic framework, such as the one we propose, we cannot calculate the marginal effects solely by translating the MM estimates through the logistic function. This is because the number of susceptible individuals in state j, day d ($S_{jd,0}$) varies depending on the treatment level, which affects the number of new cases and susceptible people every period onward. To calculate the marginal effect of interest at a particular point in time (day D since the onset of the virus), we use estimated MM parameters (shown in Tables B.2, B.1, and B.3) to simulate the new daily number of cases in each state and then aggregate the simulated new cases under the current and counterfactual levels of the targeted variable. As a result, we construct each marginal effect by differentiating the simulated number of new cases under competing scenarios. We have:

$$\Delta_D = \sum_{d=1}^{D} \left(\frac{S_{jd,0}}{1 + exp\left(-\zeta_{jd}\left(\widehat{\vec{\beta}_X}, \widehat{\vec{\beta}_\Gamma}, \widehat{\vec{\beta}_D}, \widehat{\vec{\beta}_G}, \widehat{\vec{\beta}_T}, \widehat{\vec{\beta}_\Sigma}, \widehat{\beta_R} | X, \Gamma, D, G_T, T, \Sigma, R\right)\right)} \right) - \sum_{d=1}^{D} \left(\frac{S_{jd,\Delta}}{1 + exp\left(-\zeta_{jd}\left(\widehat{\vec{\beta}_X}, \widehat{\vec{\beta}_\Gamma}, \widehat{\vec{\beta}_D}, \widehat{\vec{\beta}_G}, \widehat{\vec{\beta}_T}, \widehat{\vec{\beta}_\Sigma}, \widehat{\beta_R} | X, \Gamma, D, G_{T+\Delta}, T+\Delta, \Sigma, R\right)\right)} \right)$$

Note that $S_{jd,0}$ is a function of lagged number of new cases under the current level of treatment, and is different from its counterpart in the second term which is affected by a change in treatment level (Δ) and affects the whole trajectory of the number of new cases.

In our simulation-based approach, we assume the social distancing restrictions and the day on which the first case occurs in each state are exogenous. Next, we simulate the daily spread of the virus under two scenarios: current and counterfactual level of treatment variable. The standard errors are bootstrapped using 100 iterations, where in each iteration we draw the parameters from the multivariate normal distribution using the mean and standard errors provided by the MM estimation ($\widehat{\vec{\beta}_X}, \widehat{\vec{\beta}_D}, \widehat{\vec{\beta}_G}, \widehat{\vec{\beta}_T}, \widehat{\vec{\beta}_\Sigma}, \widehat{\beta_R}$).

Since the current policy is also simulated, we can compare the goodness of fit—resulting from the MM estimator—by comparing observed data of the spread of the virus and simulated current policy. Figure 2.11 depicts the continuous trajectory of total cases, for the U.S. and the four census regions.

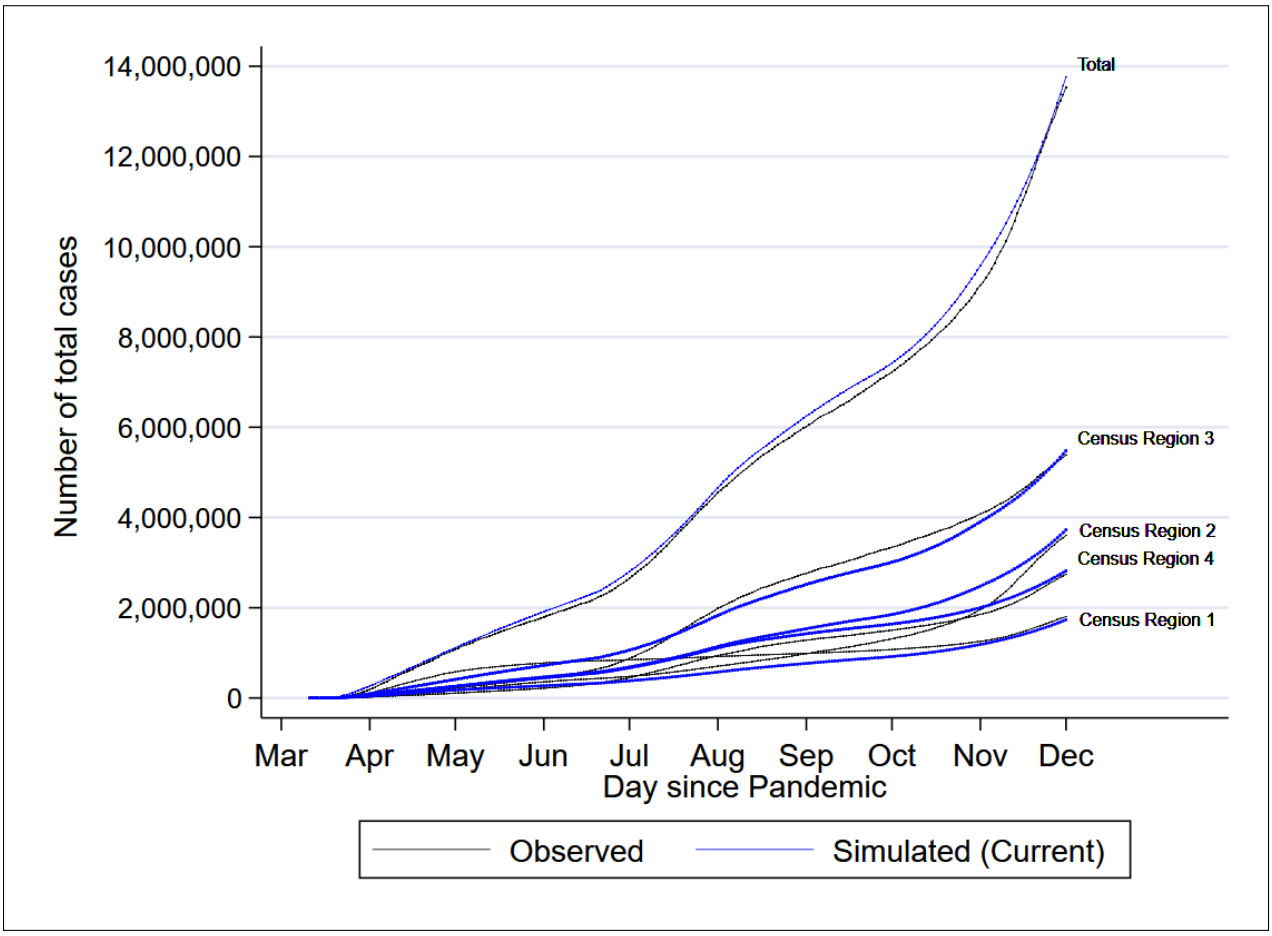


Figure 2.11: Total COVID-19 cases (observed and simulated)

The observed and simulated cases of COVID-19 are shown in the graph for the whole United States and different census regions, as suggested by the U.S. Census Bureau.

As mentioned above, the next step is to show the marginal effects by aggregating over the differences in new cases under two scenarios. Table 2.14 summarizes the simulated marginal effects of extensive and intensive margins at the end of November for three groups: Union members, covered nonmembers, and all employees. Extensive margins are based on raising employees by 1,000 for each group of employees. For intensive margins, we simulate the

counterfactual case in which we increase weekly hours of work by one hour for each group. Table 2.14 also compares the marginal effects with other specifications that lacked some or all the treatment controls.

The marginal effects in our model of choice (model 5) are robust to the exclusion of different treatment controls. Model 1 does not contain any of the treatments (extensive and intensive for three groups of employees). It is included to show the goodness of fit in the absence of treatments. If treatment variables do not contribute to prediction and only reflect the treatment effects, we expect the mean squares of errors (MSE) to be the same for all the models in Table 2.14. However, as shown in the Table 2.14 models with treatments (2-5) result in slightly smaller MSE, showing the treatments are marginally involved in the prediction, with no additional predictive power as we increase the number of treatment variables from model 2 to 5. A comparison between models with treatments shows the orthogonality between different treatments. As it is shown in Table 2.14 exclusion of different treatments from model 5 (our model of choice) does not affect the marginal effects significantly. By excluding hours of work for all employed workers (model 4), the intensive margin for the union members decreases slightly (from -211,184 to -175,099 for union members) while the sign is robust to this exclusion.

To investigate the robustness of the extensive margin, we can compare models 2 and 3. Again, the marginal effects are robust to the exclusion of employment level for all employees. There is no significant difference in the marginal effects between different models; however, there is a clear distinction between the employed and organized labor force. There is a consistency between union members and covered nonmember individuals in both extensive and intensive margins, where both have a decreasing effect on the number of COVID-19 cases. This implies we sufficiently controlled for the characteristics governing the union membership choice to such a degree that conditional on these sets of controls (such as socioeconomic characteristics and occupation shares) those who made the choice to become member of union and those who opted out (covered nonmembers) affect the spread of the virus almost

similarly. In clear contrast, both margins are positive for the employed group. Hence, while employment per se increases the spread of COVID-19, a unionized labor force can reverse this effect.

As shown in the last column of the Table 2.14, while the effect of change in the extensive margin is comparable in absolute value for the three groups of employees, the effects for intensive margins (hours of work in the previous week) are highly heterogenous comparing members, covered nonmembers, and all employed. This is mainly a by-product of the difference between the size of each group. While the intensive margin refers to one more hour of work per week, there are significantly more employed individuals than union members and much more than covered nonmembers. As a result, one more hour of work for all employees affects the COVID-19 cases more severely than one more hour of work for union members, which are much smaller in size.

Figure 2.12 depicts the daily marginal effects resulted from simulations based on model 5 and shows the dynamic change in the total number of COVID-19 cases due to each treatment. It includes marginal effects of increasing each group's size by 1,000 employees (top graphs) and increasing hours of work for each group in the previous week by one hour (bottom graphs). It shows a clear distinction in the patterns of the marginal effects between the organized labor force (union members and covered nonmembers) and all the employees. Figure 2.12 also documents more pronounced effects later in the year as the number of cases has increased dramatically.

As discussed previously, include covered individuals in the model and utilize 2019 labor force data (instead of current 2020 data) address the union membership endogeneity bias and simultaneity bias, respectively. Table 2.15 documents the results of excluding each of the two measures all based on model 5 of the Table 2.14. The first column in Table 2.15 excludes both measures. That is, it utilizes 2020 labor force data and does not include covered nonmember employees. As a result, it is prone to both biases. The second and third columns add covered nonmember employees and employs 2019 labor force data, respectively.

Table 2.14: Simulated marginal effects

Model			1	2	3	4	5
Simulated Marginal Effect $\vec{T}$	Extensive Margin (Level)	Member		-136	-128	-146	-110
				(9)	(6)	(9)	(8)
		Covered		-182	-288	-276	-256
				(20)	(21)	(32)	(20)
		Employed			549	544	507
					(17)	(17)	(22)
	Intensive Margin (Hours)	Member				-175,099	-211,184
						(10,901)	(17,171)
		Covered				-44,568	-42,432
						(10,515)	(7,923)
		Employed					1,589,455
							(88,354)
MSE			667,428	574,950	560,164	557,691	560,270
MM criterion			0	0	0	0	0
# of Observations			13,944	13,944	13,944	13,944	13,944

This table compares the marginal effects for both extensive (employment level) and intensive (hours of work during the past week) margins for five specifications differing in the number of treatment controls. As a result, each cell in the table is individually simulated. These simulations are repeated for each cell to construct the bootstrapped standard errors of the effects. An extensive margin is calculated by increasing the number of employees in each group by 1,000. For example, -110 in model 5 for members means that by increasing union members by 1,000, the accumulated effect from the beginning of the pandemic until the end of November leads to 110 fewer cases. Intensive margin, on the other hand, is the effect of one hour more of weekly work. Hence, while extensive margins are homogenized between three groups, intensive margins can be affected by each group's size. All specifications include controls for $\vec{X}, \vec{\Gamma}, \vec{D}, \vec{G}, \vec{\Sigma}$, and R. Socioeconomic ($\vec{X}$) and share ($\vec{\Sigma}$) variables include characteristics associated with each treatment group (union members, covered non-members, and all employees) depending on the specification. For specification 1 with no treatment, we use $\vec{X}$ and $\vec{\Sigma}$ for all employees. Bootstrapped standard errors are in parentheses.

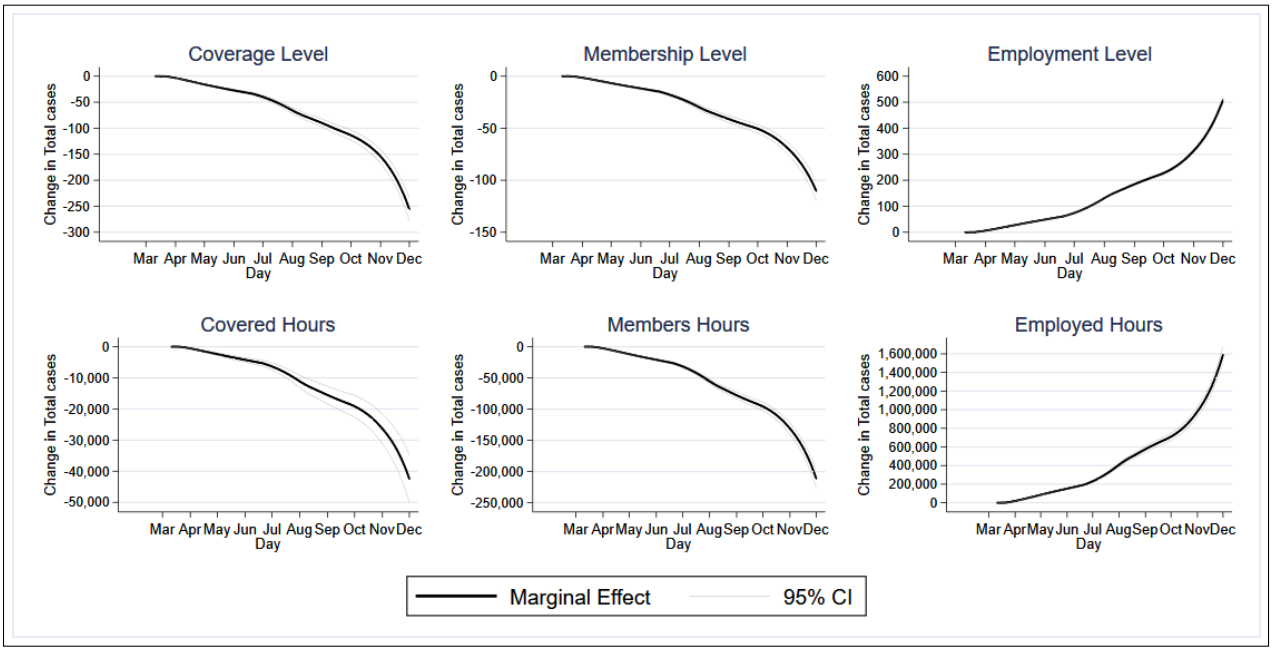


Figure 2.12: Dynamic change in total number of COVID-19 cases due to change in each treatment.

Each graph shows the daily change in the total number of cases due to a particular treatment. Graphs in the first row show the effect of increasing the number of employees in each group: covered nonmembers(left), union members (middle), all employees (right). The second row includes graphs showing the effect of increasing weekly hours of work by one hour for each group, following the same order as the graphs in the first row.

The last column includes both of the measures and is the same as model 5 in Table 2.14.

Comparing the third and the forth column suggests that inclusion of the covered employees does not affect the marginal effects for union members significantly and reflects the sufficient controls in our chosen specification that prevents the contamination of the marginal effects due to the union member different characteristics and behavior. This is stressed out again in the last column itself where the marginal effects for the union members and covered nonmembers are in the same direction with twice as large of the effect for the covered group. Comparing second and the last column shows the simultaneity bias effect. Using 2020 labor force data in the second column flips the direction of the extensive margin effects for both covered employees and all employed individuals. While the direction is the same for the union members, it changes significantly in absolute value (from -168 in the second

Table 2.15: Effect of controlling for the simultaneity and selection bias.

			1	2	3	4
Simultaneity bias control	(Lagging)				✓	✓
Selection Bias control	(Covered)			✓		✓
Simulated Marginal Effect	Extensive Margin (Level)	Member	-160	-168	-92	-110
			(12)	(14)	(5)	(12)
		Covered		149		-256
				(38)		(29)
		Employed	-323	-296	466	507
			(15)	(20)	(36)	(21)
	Intensive Margin (Hours)	Member	-349,745	-374,984	-171,268	-211,171
			(20,348)	(16,901)	(27,152)	(15,985)
		Covered		-74,766		-42,432
				(8,612)		(10,239)
		Employed	2,843,070	2,882,600	1,232,103	1,589,448
			(126,982)	(178,874)	(105,972)	(88,004)
MSE			522,055	516,830	589,213	560,270
MM criterion			0	0	0	0
# of Observations			13,944	13,944	13,944	13,944

This table shows the effect of controlling for each endogeneity bias based on specification 5 of the table 2.14. Column 1 includes none of the two controls. That is it employs 2020 labor force data without covered nonmembers group. Column 2 adds the covered nonmembers. The first two columns utilize 2020 labor force data, which is prone to simultaneity bias due to the massive impact of the COVID-19 on employment. Columns 3 and 4, on the other hand, employ 2019 labor force data to avoid the simultaneity bias, with model 4 adding the covered nonmembers to model 3. Model in column 4 is the same as specification 5 in the table 2.14. All models include controls for $\vec{X}, \vec{\Gamma}, \vec{D}, \vec{G}, \vec{\Sigma}$, and R. Socioeconomic ($\vec{X}$) and share ($\vec{\Sigma}$) variables include characteristics associated with each treatment group (union members, covered non-members, and all employees) depending on the model. Bootstrapped standard errors are in parentheses.

column to -110 in the last column), showing the simultaneity bias as expected due to the severe impact of the COVID-19 on the extensive margin. Also, consistent with the previous studies, non-organized labor force are affected more severely than union members. On the other hand, intensive margins do not show a significant change when we utilize lagged data of the labor force.

Our results are robust to the changes in the regression method we employ. Estimating the same specification as the one in column five of the Table 2.14 (our main specification) using OLS and median regression, we find both regressions approximately converge to the same results as MM. Specifically, all of the estimates are similar in the sign of the marginal effects.

MM estimates are smaller in absolute values. Both OLS and median regressions rely on the log transformation of the odd ratios, and hence, drop the observations with 0 odd ratios. This can happen in the early stages of pandemic in each state, where it is likely to have days with no reported new case. We can solve this issue by replacing the associated odd ratios with a small positive value. As it is discussed in section 2.6, depending on the replacing value, the results vary. As expected, OLS shows higher sensitivity to these changes than median regression. On the other hand, our MM strategy relies on the nonlinear probability model with no transformation and hence does not encounter this issue.

2.5.1 Employment Dynamics

As explained in section 2.3.2, we expect the COVID-19 labor market shock to be more substantial for non-organized labor force than unionized employees. Figure 2.13 shows the changes in the different labor force characteristics for all and organized employees. To better represent the dynamics and to be able to compare changes for both groups, we normalize both time-series by setting the value of each dimension at 100 in January. While the Employment level for all employees has reduced dramatically in April, employment among the unionized employees is statistically the same as in January. However, as shown by the hours of work in the previous week, the intensive margin of labor supply is decreased for both groups, with more substantial drop for unionized employees. Usual hours of work have increased for all employees, whereas it is decreased for unionized workers. These trends can be justified if we assume the probability of layoffs is higher for part-time employees for the unorganized labor force.

On the other hand, for unionized employees, since layoff is costly for employers, they might use early retirement to manage the workforce flow [Rones (1981)]. As a result, after being hit by the COVID-19 pandemic shock, the organized labor force tends to be younger and less educated than all employees. Note that changes in the organized labor force are not statistically significant for usual hours of work and education. Lastly, income has increased

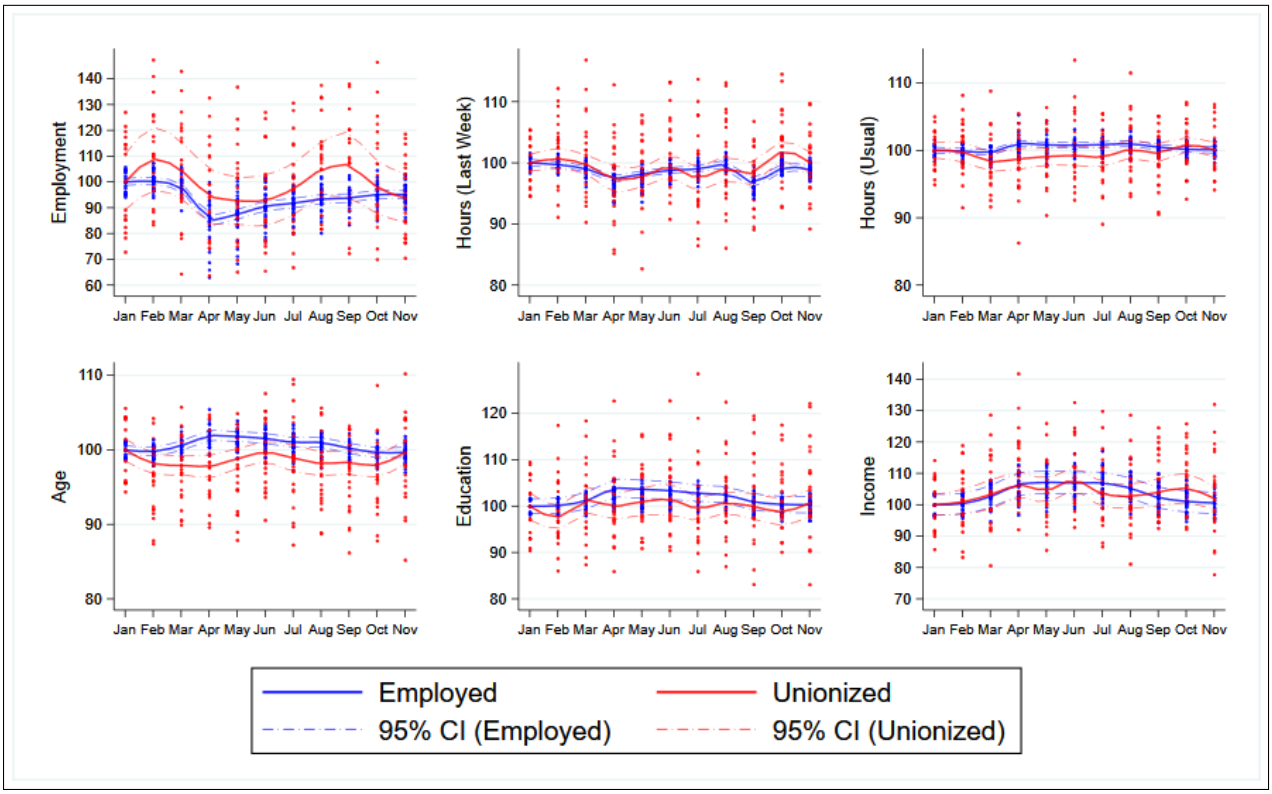


Figure 2.13: Employment dynamics in 2020

Normalized monthly changes in employment (employment level, hours in previous week, and usual hours of work) and individual characteristics (age, education, and income) for the union members and all employees in 2020 are shown in the graph.

for both groups, mostly due to the economic relief plan that took effect in April.

2.5.2 Discussion

Our results can be utilized to highlight the differences in the objective functions of government and individuals. In the case of an unpredicted pandemic such as COVID-19, the weights governments assign to people's consumption and health in the social welfare function is not necessarily correct (i.e., up to date). While federal and state governments set restrictions to control the spread of the virus, individuals are in a better position to assess their health, financial condition, and preferences regarding health and consumption. We can test this hypothesis by comparing union members (whose voice is echoed) with employed individuals not covered by a union. The results in Table 2.14 (model 5) show a negative effect of unionization on the spread of the COVID-19 virus in both extensive and intensive margins. Both of these effects are positive for all employees.

Both the monopoly face and the unions' institutional response face may contribute to these results. This study's design does not allow us to separate the effects of each of the two dimensions. However, we can associate the extensive margin (union size) with the monopoly face and the intensive margin (hours of work during the previous week) with the institutional face. Regarding the monopoly face, union members are benefited from more job safety, paid sick leave, and employer-provided health insurance, which all can protect union members from getting infected by COVID-19, mostly due to transferring the costs to the employer. Institutional face, on the other hand, facilitates the preference revelation and more optimum level of protection. Unions—due to their collective power—have more visibility and hence can employ measures closer to the collective optimum of their coalition. Our results show the hours of work in the past week negatively affects the spread of the virus for the union members, indicating a safer workplace. For all employees, this effect is positive. Both signs are consistent with the hypothesis stating there are better workplace safety measures for the union members than all employees.

If we assume that larger unions have more monopoly power, the extensive margin (union size) effect can be interpreted as the effect of the monopoly power and better contracts in unions than all employees. Lastly, likely, the gap between union members and all employees in usual hours of work, age, and education level (shown in Figure 2.13) is due to unions' monopoly power. Lastly, considering the Centers for Disease Control and Prevention (CDC) estimates based on antibody test results[18], unions' positive effect can be even more significant, suggesting our results underestimate the true effect.

2.6 Robustness Check: OLS and Median Regression Results

We can estimate the logistic models by transforming the outcome to the natural logarithm of the odds ratios:

$$y_{jd} = \frac{S_{jd}}{1 + exp(-\vec{Z}_{jd}.\vec{\beta}')} \rightarrow Log(\frac{y_{jd}}{S_{jd} - y_{jd}}) = \vec{Z}_{jd}.\vec{\beta}' \tag{2.18}$$

This transformation allows us to estimate the logistic model using the OLS estimation method. However, if the outcome $y_{jd} = 0$, the logarithm is undefined. This problem leads to the dropping of all those observations with an outcome at zero. If deletion of the outcomes is random, we do not expect to see severe changes in the estimation results comparing to the non-linear model (without transformation). However, if the treatment is non-randomly affected by these deletions, the calculated marginal effects will be biased. Our analysis shows we have 366 observations with no new cases. That is, after observing the first COVID-19 case in state j, there are days with no new cases recorded. The level of covered non-members employment is higher for those states where COVID-19 spreads more slowly in the first months after the beginning of the pandemic in the U.S. As a result, we expect the estimates and the marginal effects for the covered non-member group to be sensitive to the deletion of the observations with no new case.

[18] Suggesting that as many as ten times more Americans may be infected with coronavirus than recognized by the confirmed number of cases [The New York Times (2020b)].

The results for comparing various OLS regressions (differing in the replacement value of the zero cases) and MM based on the model 5 of the Table 2.14 are shown in Table 2.16. While dropping the outcomes with zero new cases (first column) changes the sign of the coefficients for covered non-members in intensive and extensive margins, assigning a small positive number retains the direction of the MM coefficients. Table 2.17 replicates the same exercise as in Table 2.16 with median regression. While they are less sensitive to the replacement values, they obtain larger parameters in absolute values.

Table 2.16: Sensitivity of OLS estimates to the odd ratios close to zero.

		OLS 0	OLS 1.E-03	OLS 1.E-04	OLS 1.E-05	OLS 1.E-06	MM -
Extensive Margin	Member	-0.077	-0.132	-0.135	-0.139	-0.143	-0.070
		(0.0159)	(0.0237)	(0.0275)	(0.0318)	(0.0362)	(0.0173)
(Level)	Covered	0.168	0.020	-0.061	-0.142	-0.223	-0.168
		(0.0605)	(0.0898)	(0.1045)	(0.1205)	(0.1373)	(0.0609)
	Employed	0.018	0.026	0.027	0.028	0.029	0.033
		(0.0036)	(0.0054)	(0.0062)	(0.0072)	(0.0082)	(0.0038)
Intensive Margin	Member	-0.025	-0.026	-0.026	-0.025	-0.025	-0.009
		(0.0022)	(0.0033)	(0.0038)	(0.0044)	(0.0050)	(0.0022)
(Hours)	Covered	0.000	-0.002	-0.003	-0.004	-0.004	-0.002
		(0.0010)	(0.0015)	(0.0017)	(0.0020)	(0.0023)	(0.0012)
	Employed	0.103	0.156	0.175	0.195	0.214	0.065
		(0.0108)	(0.0160)	(0.0186)	(0.0215)	(0.0244)	(0.0130)
# of Observations		13,578	13,944	13,944	13,944	13,944	13,944

Each column assigns a specific value (shown in each column header) to the outcomes with zero new cases. For the first column, when OLS assigns zero to the outcomes (keeping them unchanged), number of observations drops from 13,944 to 13,578. Assigning a small positive number can retain the direction of the MM estimates. Standard errors are in parentheses.

2.7 Conclusion

Efforts to stop the spread of COVID-19 can be seen as a public good. In other words, it is costly for agents (individuals or firms) with a particular level of effort, while other individuals can reap the benefits. As a result, the benefits of the efforts cannot be constrained to the agent who bears the costs. Thus, agents with the highest preferences for health, those with poorer health, or those who are wealthier possibly bear the costs, and other agents take a

Table 2.17: Sensitivity of median regression to the odd ratios close to zero.

		$q(50\%)$ 0	$q(50\%)$ 1.E-03	$q(50\%)$ 1.E-04	$q(50\%)$ 1.E-05	$q(50\%)$ 1.E-06	MM -
Extensive Margin	Member	-0.158	-0.152	-0.153	-0.156	-0.158	-0.070
		(0.0228)	(0.0241)	(0.0256)	(0.0269)	(0.0285)	(0.0173)
(Level)	Covered	0.017	-0.112	-0.141	-0.144	-0.163	-0.168
		(0.0849)	(0.0895)	(0.0950)	(0.0998)	(0.1058)	(0.0609)
	Employed	0.017	0.025	0.024	0.024	0.024	0.033
		(0.0056)	(0.0059)	(0.0063)	(0.0066)	(0.0070)	(0.0038)
Intensive Margin	Member	-0.040	-0.040	-0.041	-0.040	-0.041	-0.009
		(0.0032)	(0.0034)	(0.0036)	(0.0038)	(0.0040)	(0.0022)
(Hours)	Covered	-0.005	-0.006	-0.006	-0.006	-0.005	-0.002
		(0.0014)	(0.0015)	(0.0016)	(0.0017)	(0.0018)	(0.0012)
	Employed	0.143	0.164	0.168	0.168	0.170	0.065
		(0.0154)	(0.0162)	(0.0172)	(0.0181)	(0.0191)	(0.0130)
# of Observations		13,578	13,944	13,944	13,944	13,944	13,944

Each column assigns a specific value (shown in each column header) to the outcomes with zero new cases. For the first column, when median regression assigns zero to the outcomes (keeping them unchanged), number of observations dropped from 13,944 to 13,578. Assigning a small positive number can retain the direction of the MM estimates. Standard errors are in parentheses.

'free ride'. Like other examples of public good, efforts to prevent the spread of the virus are under-supplied, and to achieve the optimum level of a public good, the government or a third-party entity should intervene and provide motivation and regulations to satisfy a certain level of public good in society. This paper investigates the role of labor unions in the spread of a contagious virus such as COVID-19.

Unions may affect the work environment through the channels discussed in the literature, namely, monopoly power and information transparency. Depending on the structure of the union and the characteristics of union members, they may use their collective bargaining power to increase safety in the work environment, leading to a decrease in the spread of the virus. Furthermore, they may more efficiently transmit employees' private health information to the firm, and as a result, establish safety measures that are closer to the optimum level. Also, unionized workers have a clearer picture of the firm's profit perspective since they can negotiate terms with their employers. Hence, a more efficient outcome in terms of employment and workplace safety is expected. However, a more efficient outcome does not

necessarily mean that the virus will spread less.

Our analysis shows that ceteris paribus, a local increase in unionization level, leads to better control of the spread of COVID-19. This is despite the fact that union members have not decreased their hours of work in response to the spread of the virus. In other words, unionization has positive externalities that are reflected in a slower spread of the virus. In this study, we utilize the confirmed COVID-19 cases to measure the spread of the virus and conclude that increasing both extensive and intensive margins of labor supply for union members slows the spread of the COVID-19 virus. To establish a more robust relationship, we compare the results with those for covered non-members and all employees. We also employ lagging variables for the labor force characteristics to avoid simultaneity bias.

This study does not address the general equilibrium aspects of the change in union size. That is, labor supply (for both intensive and extensive margins) and a firm's profit are altered neither by direct changes in unionization level nor by indirect changes due to the effect of unionization on the spread of the virus. To analyze the net and longer-term effects of a change in unionization further studies with general equilibrium aspects that endogenize labor market responses are required.

Appendix A

Appendix for Chapter I

A.1 Additional tables

Below are the supplementary tables for chapter I. Tables A.1 and A.2 report OLS and DiD results between RC and other climate-related variables for the fixed scheme. Tables A.3, A.4, and A.5 report the IV result for the inner-city outer city and the difference between in and out of the city for the fixed scheme, respectively. Tables A.6 to A.10 reports the same result for the flexible scheme.

Table A.1: OLS results of the effects for multiple outcomes (Fixed Scheme)

	Annual	Winter	Spring	Summer	Autumn
Dew Point	-8.08	-6.18	-6.73	-6.99	-9.29
	(6.56)	(11.25)	(5.62)	(3.42)	(8.42)
Sea Level Pressure	-293.37	-253.91	-250.29	-262.46	-253.16
	(2394.68)	(2355.13)	(2312.36)	(2510.31)	(2493.84)
Station Pressure	218.8	196.54	217.55	241.94	257.7
	(6065.7)	(6195.98)	(6081.14)	(6973.84)	(6548.83)
Visibility	-21.33	-5.07	-28.77	-5.71	-7.08
	(451.)	(404.65)	(333.13)	(456.35)	(484.44)
Wind Speed	5.78	7.04	4.75	7.2	7.22
	(1.56)	(1.91)	(.8)	(1.98)	(2.11)
Maximum Wind Speed	23.94	38.09	10.27	42.2	38.39
	(183.48)	(207.4)	(96.18)	(224.95)	(231.06)
Gust	4.56	13.74	2.82	1.21	2.01
	(9.41)	(9.9)	(13.37)	(11.21)	(9.92)
Percipitation	-.05	-.04	-.03	-.09	-.05
	(.)	(.)	(.)	(.)	(.)
Temperature	-12.11	-16.36	-10.17	-8.19	-16.54
	(3.64)	(7.41)	(2.69)	(2.2)	(4.51)
Number of Hot Days	4.41	2.55	.	.	.95
	(.24)	(.13)	(.)	(.)	(.02)
Number of Very Hot Days	-5.06	.	-.87	-4.21	.
	(2.74)	(.)	(.21)	(1.56)	(.)
Number of Average Days	-68.13	-5.88	-26.5	-24.19	-13.48
	(149.06)	(1.22)	(25.71)	(28.3)	(5.93)

Standard errors in parentheses

Table A.2: Difference-in-Difference effects for multiple outcomes (Fixed Scheme)

	Annual	Winter	Spring	Summer	Autumn
Dew Point	-1.49	-.7	-.92	-.89	-2.01
	(.65)	(.67)	(.59)	(.54)	(.76)
Sea Level Pressure	16.78	31.45	29.3	29.68	23.75
	(16.17)	(15.17)	(15.96)	(16.17)	(16.16)
Station Pressure	73.94	108.06	67.13	62.62	77.31
	(21.69)	(21.16)	(21.88)	(24.73)	(23.72)
Visibility	-41.14	-28.37	-33.81	-37.43	-41.27
	(7.91)	(6.72)	(7.43)	(7.84)	(7.91)
Wind Speed	.	.09	.14	.17	.05
	(.)	(.09)	(.09)	(.08)	(.08)
Maximum Wind Speed	3.46	4.6	2.91	3.95	3.69
	(1.54)	(1.59)	(1.4)	(1.52)	(1.49)
Gust	-.31	.47	-.75	.35	.47
	(.93)	(1.07)	(1.11)	(1.06)	(.98)
Percipitation	-.01	-.02	-.02	-.01	-.01
	(.01)	(.01)	(.01)	(.01)	(.01)
Temperature	.59	.12	.52	.59	.12
	(.27)	(.35)	(.29)	(.27)	(.38)
Number of Hot Days	-.01	-.02	.	.	.02
	(.16)	(.12)	(.)	(.)	(.05)
Number of Very Hot Days	.72	.	.24	.49	.
	(.67)	(.)	(.19)	(.49)	(.)
Number of Average Days	7.57	.31	2.63	3.79	.84
	(2.59)	(.58)	(1.02)	(1.46)	(.79)

Standard errors in parentheses
Causal Effect of Residential Compactness on multiple environmental dimensions for the center of middle size MSAs.

Table A.3: IV effects for multiple outcomes in urban area (Fixed Scheme)

	Annual	Winter	Spring	Summer	Autumn
Dew Point	-3.47	15.22	-34.09	-13.55	-23.06
	(11.35)	(14.67)	(10.21)	(8.05)	(13.3)
	39.66	41.73	43.67	42.36	36.23
Sea Level Pressure	-870.92	-1035.24	-849.04	-923.61	-1023.09
	(276.65)	(274.83)	(275.36)	(278.02)	(280.93)
	16.81	17.08	15.90	18.36	17.38
Station Pressure	-1036.73	-1896.37	-1124.61	-936.87	-1042.03
	(370.62)	(404.4)	(362.33)	(391.38)	(394.08)
	32.15	25.69	34.93	33.91	30.13
Visibility	-493.04	-442.23	-420.78	-471.35	-452.19
	(142.49)	(130.66)	(121.23)	(139.68)	(143.82)
	6.33	8.28	6.91	7.83	7.87
Wind Speed	2.86	5.59	6.76	-1.97	-.63
	(8.29)	(8.48)	(5.82)	(8.8)	(9.07)
	6.86	11.44	8.21	10.22	10.38
Maximum Wind Speed	-136.8	-148.43	-93.97	-150.05	-162.33
	(93.44)	(91.86)	(68.03)	(96.12)	(98.33)
	4.70	8.94	4.39	8.65	8.17
Gust	215.13	159.38	284.18	261.54	132.54
	(17.66)	(18.92)	(22.19)	(19.24)	(16.96)
	12.45	10.95	9.86	11.79	17.88
Percipitation	.61	1.39	.99	.19	.13
	(.09)	(.14)	(.14)	(.15)	(.08)
	9.85	7.10	6.71	4.44	11.37
Temperature	-14.35	14.37	-29.6	-21.71	-24.27
	(4.53)	(5.83)	(3.55)	(5.49)	(5.77)
	186.73	235.62	229.69	65.90	139.54
Number of Hot Days	1.61	1.11	.	.	.75
	(2.75)	(2.16)	(.)	(.)	(1.07)
	17.57	13.22	0.00	0.00	3.81
Number of Very Hot Days	-15.05	.	-4.72	-10.3	.
	(5.24)	(.)	(1.76)	(4.14)	(.01)
	96.63	0.00	59.24	86.48	1.24
Number of Normal Days	-215.51	16.18	-108.3	-90.73	-32.88
	(32.06)	(5.)	(14.66)	(18.5)	(7.34)
	150.60	38.01	121.42	78.20	109.83

Standard errors in parentheses and third row represents F-statistic for the 1st stage.
Causal Effect of Residential Compactness on multiple environmental dimensions for the center of middle size MSAs.

Table A.4: IV effects for multiple outcomes in surrounding area (Fixed Scheme)

	Annual	Winter	Spring	Summer	Autumn
Dew Point	-44.64	-26.48	-57.45	-37.1	-57.58
	(9.27)	(11.25)	(8.5)	(6.6)	(10.68)
	22.17	26.22	25.59	25.78	20.55
Sea Level Pressure	-1197.23	-1044.76	-1044.99	-1077.67	-1224.74
	(237.43)	(233.98)	(238.66)	(237.69)	(239.99)
	14.84	15.12	14.77	15.06	14.67
Station Pressure	-469.89	-597.33	-450.14	-487.45	-407.39
	(223.59)	(235.91)	(220.84)	(228.63)	(233.88)
	21.29	17.98	23.77	23.60	19.68
Visibility	-1041.94	-920.54	-959.54	-1029.64	-1005.35
	(122.87)	(114.02)	(119.24)	(123.01)	(122.02)
	7.88	8.11	8.29	8.34	8.08
Wind Speed	.	3.7	1.28	-4.12	-1.88
	(.)	(1.76)	(2.15)	(2.39)	(2.23)
	0.00	2.45	3.05	2.54	1.81
Maximum Wind Speed	-100.29	-91.67	-126.12	-127.98	-122.71
	(20.64)	(26.37)	(27.92)	(32.64)	(33.72)
	17.35	10.43	7.74	7.75	6.75
Gust	177.79	153.72	224.64	210.65	133.08
	(13.69)	(14.11)	(16.39)	(15.11)	(12.62)
	9.55	7.96	9.25	10.31	12.61
Percipitation	.33	.67	.46	.07	.07
	(.05)	(.08)	(.08)	(.07)	(.05)
	14.59	10.60	9.49	8.93	12.33
Temperature	-5.84	12.77	-23.11	-12.46	-13.89
	(2.58)	(4.01)	(3.12)	(2.78)	(3.45)
	307.16	264.14	173.56	119.68	201.81
Number of Hot Days	.54	.72	-.01	.	.2
	(1.74)	(1.35)	(.02)	(.)	(.72)
	19.02	14.96	0.93	0.00	3.95
Number of Very Hot Days	-3.66	-.18	-.19	-3.25	-.04
	(3.4)	(.08)	(.68)	(2.96)	(.05)
	5.22	1.84	3.65	5.23	0.96
Number of Normal Days	-18.55	25.11	-43.14	-1.91	1.4
	(21.02)	(3.84)	(9.87)	(12.72)	(4.52)
	196.81	39.17	148.95	87.83	153.06

Standard errors in parentheses and third row represents F-statistic for the 1st stage.

Table A.5: Causal effect of residential compactness on difference between urban and surrounding area for multiple outcomes (Fixed Scheme)

	Annual	Winter	Spring	Summer	Autumn
Dew Point	48.43	47.32	40.95	31.53	48.5
	(14.69)	(17.72)	(13.1)	(10.09)	(16.06)
	14.11	16.07	14.11	16.07	15.79
Sea Level Pressure	473.7	104.9	285.33	295.54	411.64
	(383.56)	(379.37)	(384.3)	(386.04)	(390.07)
	256.00	12.80	11.25	12.58	12.25
Station Pressure	-589.24	-1288.44	-684.4	-398.68	-560.89
	(338.22)	(352.89)	(330.37)	(371.72)	(370.2)
	24.69	21.28	23.75	23.95	23.90
Visibility	738.64	651.74	706.42	715.24	715.62
	(200.98)	(184.57)	(179.51)	(198.25)	(203.57)
	7.05	8.42	8.56	8.44	8.06
Wind Speed	2.86	1.39	5.8	4.03	.92
	(8.29)	(8.74)	(6.44)	(9.35)	(9.54)
	6.86	10.55	6.64	9.48	9.02
Maximum Wind Speed	38.72	21.19	59.07	19.35	-3.46
	(95.88)	(96.72)	(75.11)	(105.21)	(106.88)
	5.56	9.25	6.03	7.94	7.52
Gust	-13.09	-31.16	-5.2	-18.16	-37.89
	(15.72)	(16.98)	(19.62)	(16.83)	(15.61)
	8.11	7.78	6.12	6.93	9.60
Percipitation	.16	.49	.33	.05	.04
	(.09)	(.12)	(.13)	(.15)	(.07)
	5.73	2.84	3.92	3.14	7.00
Temperature	-10.75	-3.93	-5.11	-8.59	-9.35
	(4.81)	(3.32)	(3.23)	(5.24)	(4.52)
	13.89	25.06	21.70	18.10	15.73
Number of Hot Days	.77	.05	.01	.	.45
	(1.57)	(1.34)	(.03)	(.)	(.62)
	5.55	4.80	4.11	0.00	1.15
Number of Very Hot Days	-8.17	.16	-4.21	-4.3	.07
	(4.29)	(.09)	(1.6)	(3.46)	(.07)
	133.34	1.64	64.71	116.32	0.86
Number of Normal Days	-195.29	-20.2	-50.91	-89.29	-34.79
	(33.59)	(3.93)	(13.05)	(18.7)	(6.19)
	16.53	7.46	15.00	17.85	20.48

Standard errors in parentheses and third row represents F-statistic for the 1st stage.

Table A.6: OLS results of the effects for multiple outcomes (Flexible Scheme)

	Annual	Winter	Spring	Summer	Autumn
Dew Point	-5.08	-1.37	-5.42	-5.8	-5.2
	(7.01)	(12.04)	(6.17)	(3.78)	(8.97)
Sea Level Pressure	-303.22	-270.44	-255.69	-260.38	-253.33
	(2966.02)	(2901.07)	(2914.64)	(3089.1)	(3095.39)
Station Pressure	167.98	163.8	174.14	200.79	217.
	(6048.96)	(6146.75)	(6012.92)	(6895.69)	(6539.84)
Visibility	6.87	26.38	3.43	29.53	26.26
	(541.87)	(543.35)	(436.29)	(599.16)	(618.25)
Wind Speed	4.7	5.61	3.86	6.06	6.06
	(1.74)	(2.04)	(.91)	(2.14)	(2.32)
Maximum Wind Speed	15.1	28.6	3.07	28.74	27.75
	(200.1)	(223.57)	(104.95)	(238.89)	(252.51)
Gust	6.45	16.81	4.04	-.44	6.03
	(9.88)	(10.46)	(13.94)	(11.72)	(10.25)
Percipitation	-.05	-.05	-.04	-.07	-.06
	(.)	(.)	(.)	(.)	(.)
Temperature	-8.1	-11.16	-6.67	-5.69	-12.23
	(3.72)	(7.47)	(2.72)	(2.27)	(4.48)
Number of Hot Days	3.97	2.18	.	.	.88
	(.26)	(.14)	(.)	(.)	(.02)
Number of Very Hot Days	-2.83	.	-.34	-2.54	.
	(1.84)	(.)	(.13)	(1.07)	(.)
Number of Average Days	-44.75	-3.05	-18.32	-18.79	-7.38
	(149.97)	(.95)	(25.96)	(29.71)	(4.99)

Standard errors in parentheses

Table A.7: Difference-in-Difference effects for multiple outcomes(Flexible Scheme)

	Annual	Winter	Spring	Summer	Autumn
Dew Point	-1.49	-.7	-.92	-.89	-2.01
	(.65)	(.67)	(.59)	(.54)	(.76)
Sea Level Pressure	16.78	31.45	29.3	29.68	23.75
	(16.17)	(15.17)	(15.96)	(16.17)	(16.16)
Station Pressure	73.94	108.06	67.13	62.62	77.31
	(21.69)	(21.16)	(21.88)	(24.73)	(23.72)
Visibility	-41.14	-28.37	-33.81	-37.43	-41.27
	(7.91)	(6.72)	(7.43)	(7.84)	(7.91)
Wind Speed	.	.09	.14	.17	.05
	(.)	(.09)	(.09)	(.08)	(.08)
Maximum Wind Speed	3.46	4.6	2.91	3.95	3.69
	(1.54)	(1.59)	(1.4)	(1.52)	(1.49)
Gust	-.31	.47	-.75	.35	.47
	(.93)	(1.07)	(1.11)	(1.06)	(.98)
Percipitation	-.01	-.02	-.02	-.01	-.01
	(.01)	(.01)	(.01)	(.01)	(.01)
Temperature	.59	.12	.52	.59	.12
	(.27)	(.35)	(.29)	(.27)	(.38)
Number of Hot Days	-.01	-.02	.	.	.02
	(.16)	(.12)	(.)	(.)	(.05)
Number of Very Hot Days	.72	.	.24	.49	.
	(.67)	(.)	(.19)	(.49)	(.)
Number of Average Days	7.57	.31	2.63	3.79	.84
	(2.59)	(.58)	(1.02)	(1.46)	(.79)

Standard errors in parentheses
Causal Effect of Residential Compactness on multiple environmental dimensions for the center of middle size MSAs.

Table A.8: Instrumental Variable estimates for multiple outcomes (Flexible Scheme)

	Annual	Winter	Spring	Summer	Autumn
Dew Point	-9.1	9.03	-37.07	-14.14	-29.78
	(11.17)	(14.42)	(10.07)	(7.96)	(12.94)
	40.63	43.22	46.46	45.18	38.70
Sea Level Pressure	-819.84	-942.01	-806.35	-853.65	-909.08
	(310.57)	(306.55)	(307.33)	(312.92)	(313.79)
	13.39	13.90	13.74	14.59	14.42
Station Pressure	-1347.72	-2001.28	-1385.18	-1193.89	-1336.79
	(352.71)	(384.5)	(346.1)	(371.58)	(373.75)
	33.19	26.52	35.63	35.14	31.87
Visibility	-571.28	-517.13	-497.21	-524.72	-536.08
	(151.15)	(143.73)	(136.75)	(151.55)	(153.93)
	5.69	8.64	5.38	8.39	8.44
Wind Speed	-3.88	.39	-.82	-9.4	-7.55
	(8.41)	(8.51)	(5.95)	(8.91)	(9.13)
	6.92	11.04	8.19	9.76	10.65
Maximum Wind Speed	-179.11	-179.93	-140.83	-198.34	-204.36
	(94.)	(93.45)	(68.31)	(97.11)	(99.43)
	4.72	8.03	4.63	7.85	8.07
Gust	206.11	154.12	274.66	249.61	128.38
	(17.31)	(18.67)	(21.58)	(18.73)	(16.66)
	12.64	11.33	10.67	12.32	17.83
Percipitation	.63	1.36	1.	.26	.12
	(.09)	(.14)	(.13)	(.14)	(.08)
	8.80	6.81	6.54	4.43	9.85
Temperature	-13.34	14.72	-26.23	-22.99	-23.39
	(4.54)	(5.72)	(3.51)	(5.49)	(5.77)
	173.12	226.27	219.26	62.83	126.96
Number of Hot Days	1.05	.59	-.01	.	.7
	(2.77)	(2.17)	(.03)	(.)	(1.08)
	17.51	13.56	0.91	0.00	3.81
Number of Very Hot Days	-12.34	.	-2.49	-9.82	-.03
	(4.86)	(.)	(1.41)	(3.91)	(.01)
	63.59	0.00	50.09	56.40	2.38
Number of Normal Days	-184.08	17.05	-95.8	-83.07	-23.86
	(31.02)	(4.37)	(14.18)	(18.39)	(6.82)
	147.69	35.26	121.08	76.64	97.02

Standard errors in parentheses and third row represents F-statistic for the 1st stage.
Causal effect of Residential Compactness on multiple environmental diemensions for the center of middle size MSAs.

Table A.9: Analysis of the sensitivity of the rural area to the Instrument for multiple outcomes(Flexible Scheme)

	Annual	Winter	Spring	Summer	Autumn
Dew Point	-47.8	-31.32	-56.91	-40.33	-62.54
	(9.3)	(11.53)	(8.5)	(6.72)	(10.83)
	20.03	23.09	22.73	22.19	18.39
Sea Level Pressure	-876.73	-776.13	-763.01	-777.01	-903.35
	(221.36)	(220.6)	(223.67)	(223.23)	(223.29)
	15.57	15.43	15.37	15.58	15.42
Station Pressure	-361.26	-481.67	-315.09	-364.71	-341.59
	(220.82)	(231.04)	(220.27)	(228.16)	(229.63)
	23.76	20.65	25.84	25.71	22.38
Visibility	-973.52	-824.63	-920.88	-1001.85	-939.99
	(116.94)	(109.63)	(116.46)	(119.98)	(116.5)
	8.43	8.42	8.43	8.53	8.45
Wind Speed	.	4.55	2.03	-3.23	-2.27
	(.)	(1.78)	(1.73)	(1.96)	(2.15)
	0.00	2.23	3.48	3.03	1.65
Maximum Wind Speed	-73.43	-55.89	-104.1	-146.94	-110.27
	(18.96)	(25.13)	(23.52)	(39.82)	(32.7)
	19.26	10.94	10.84	5.37	7.38
Gust	147.53	127.15	183.42	167.11	104.29
	(12.89)	(13.51)	(15.29)	(14.04)	(12.17)
	10.50	8.34	9.89	10.71	13.24
Percipitation	.35	.63	.5	.12	.1
	(.05)	(.08)	(.08)	(.08)	(.05)
	13.92	10.18	8.85	8.26	11.17
Temperature	-7.99	10.51	-17.53	-7.89	-15.58
	(2.5)	(3.87)	(2.72)	(2.67)	(3.43)
	300.29	261.74	208.96	121.38	188.03
Number of Hot Days	.57	.69	.	.	.21
	(1.73)	(1.33)	(.02)	(.)	(.72)
	18.42	14.35	0.90	0.00	3.84
Number of Very Hot Days	-2.36	-.13	.14	-2.36	.
	(2.87)	(.07)	(.51)	(2.52)	(.)
	6.29	1.81	4.27	6.30	0.00
Number of Normal Days	-22.21	18.26	-37.56	-1.26	-1.66
	(20.75)	(3.34)	(9.46)	(12.47)	(4.58)
	180.99	32.64	147.93	88.67	112.13

Standard errors in parentheses and third row represents F-statistic for the 1st stage.

Table A.10: Causal effect of residential compactness on difference between urban and surrounding area for multiple outcomes

	Annual	Winter	Spring	Summer	Autumn
Dew Point	48.44	49.45	38.72	36.25	49.92
	(14.58)	(18.04)	(13.11)	(10.24)	(16.14)
	15.01	16.72	15.80	17.22	16.99
Sea Level Pressure	.49	-216.42	-107.42	-119.32	11.07
	(370.35)	(363.8)	(366.18)	(373.49)	(370.66)
	256.00	13.71	12.45	13.08	13.31
Station Pressure	-1035.32	-1525.96	-1099.47	-822.01	-959.3
	(334.12)	(343.07)	(319.19)	(360.86)	(359.62)
	24.63	21.91	24.60	24.39	24.71
Visibility	568.97	457.67	595.79	643.2	553.66
	(199.23)	(193.42)	(188.32)	(204.78)	(206.24)
	6.68	7.65	6.93	8.19	7.91
Wind Speed	-3.88	-4.82	-2.75	-4.28	-5.75
	(8.41)	(8.82)	(6.31)	(9.25)	(9.43)
	6.92	10.09	7.17	9.60	9.87
Maximum Wind Speed	-36.59	-52.01	-17.09	6.34	-63.12
	(95.81)	(97.96)	(72.3)	(111.76)	(105.91)
	5.56	8.42	6.81	6.85	7.92
Gust	21.35	2.03	44.2	31.79	-1.73
	(15.29)	(16.38)	(18.97)	(16.87)	(15.44)
	8.92	9.45	8.37	8.25	10.47
Percipitation	.15	.51	.27	.07	.
	(.09)	(.12)	(.13)	(.14)	(.08)
	5.75	2.94	4.26	3.41	6.64
Temperature	-6.69	-.84	-9.78	-16.35	-5.81
	(4.8)	(3.22)	(3.02)	(5.3)	(4.56)
	12.60	25.06	29.31	19.30	14.07
Number of Hot Days	.18	-.42	-.01	.	.41
	(1.62)	(1.39)	(.04)	(.)	(.63)
	6.56	5.88	2.50	0.00	1.24
Number of Very Hot Days	-7.36	.09	-2.44	-4.97	-.03
	(3.52)	(.07)	(1.2)	(2.93)	(.01)
	104.13	1.65	61.20	87.85	2.38
Number of Normal Days	-152.3	-9.58	-44.28	-78.5	-20.45
	(31.24)	(3.23)	(12.12)	(18.25)	(5.21)
	18.41	11.29	15.85	21.08	23.73

Standard errors in parentheses and third row represents F-statistic for the 1st stage.

A.2 Additional Figures

Below are the supplementary figures for chapter I. Figure A.1 demonstrate Choropleth map for temperature for 1974 and 2012. Similarly, figures A.2 shows the change in visibility and figures A.3 illustrates the change between 1974 to 2012 for Dew Point.

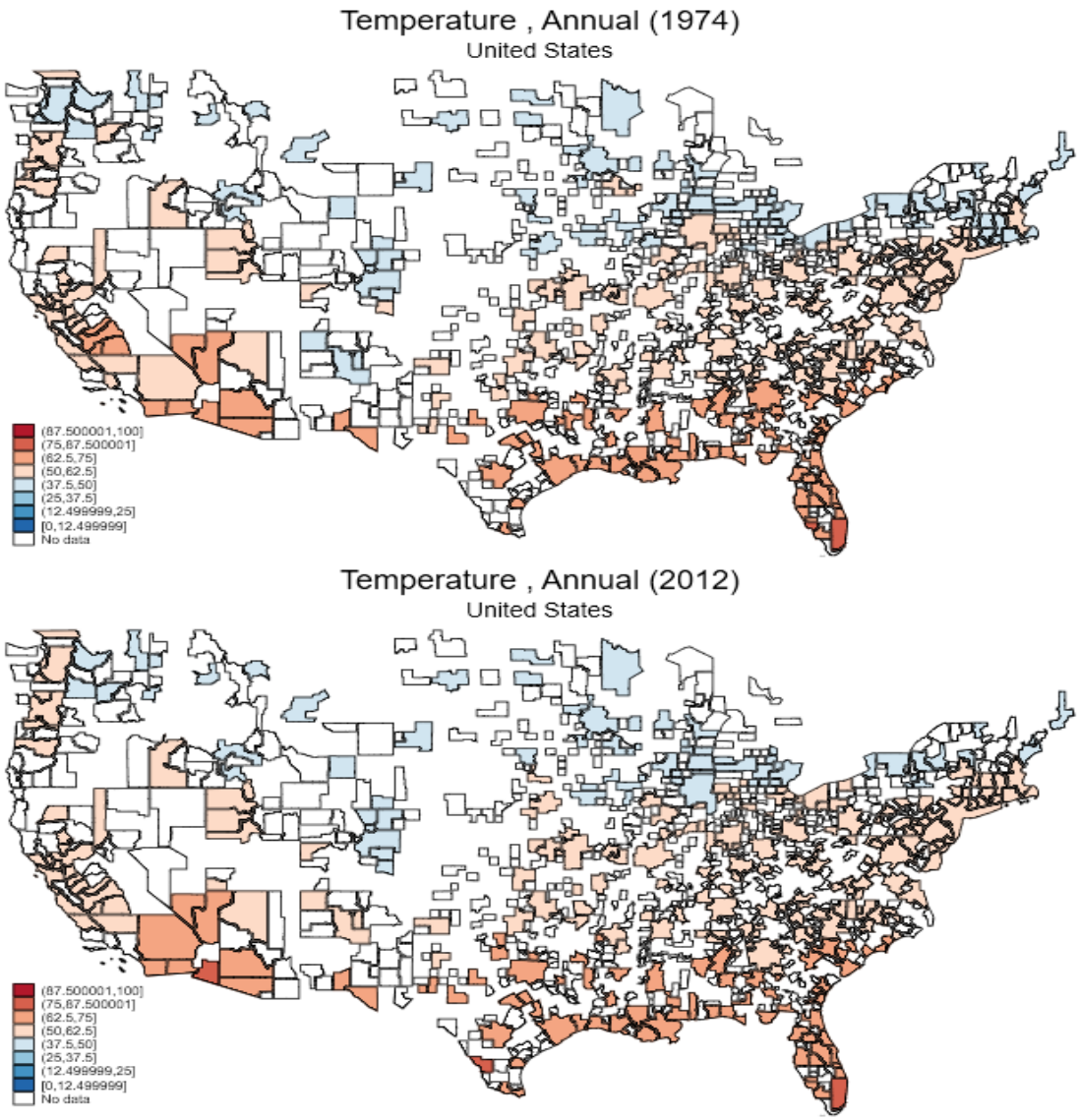


Figure A.1: GSOD data of Annual Temperature

Source: Author's Calculation using ArcGIS and GSOD

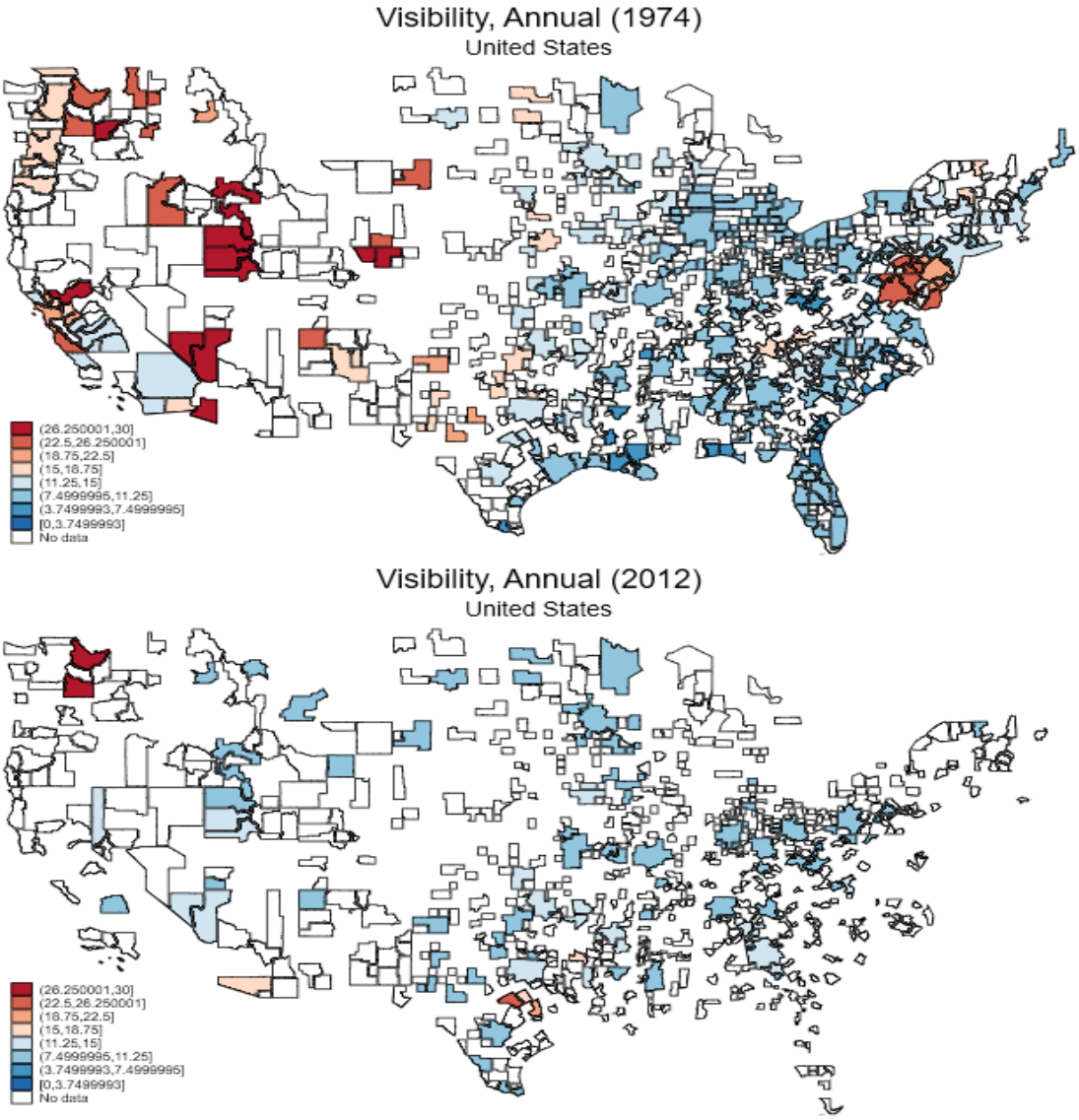


Figure A.2: GSOD data of Visibility

Source: Author's Calculation using ArcGIS and GSOD

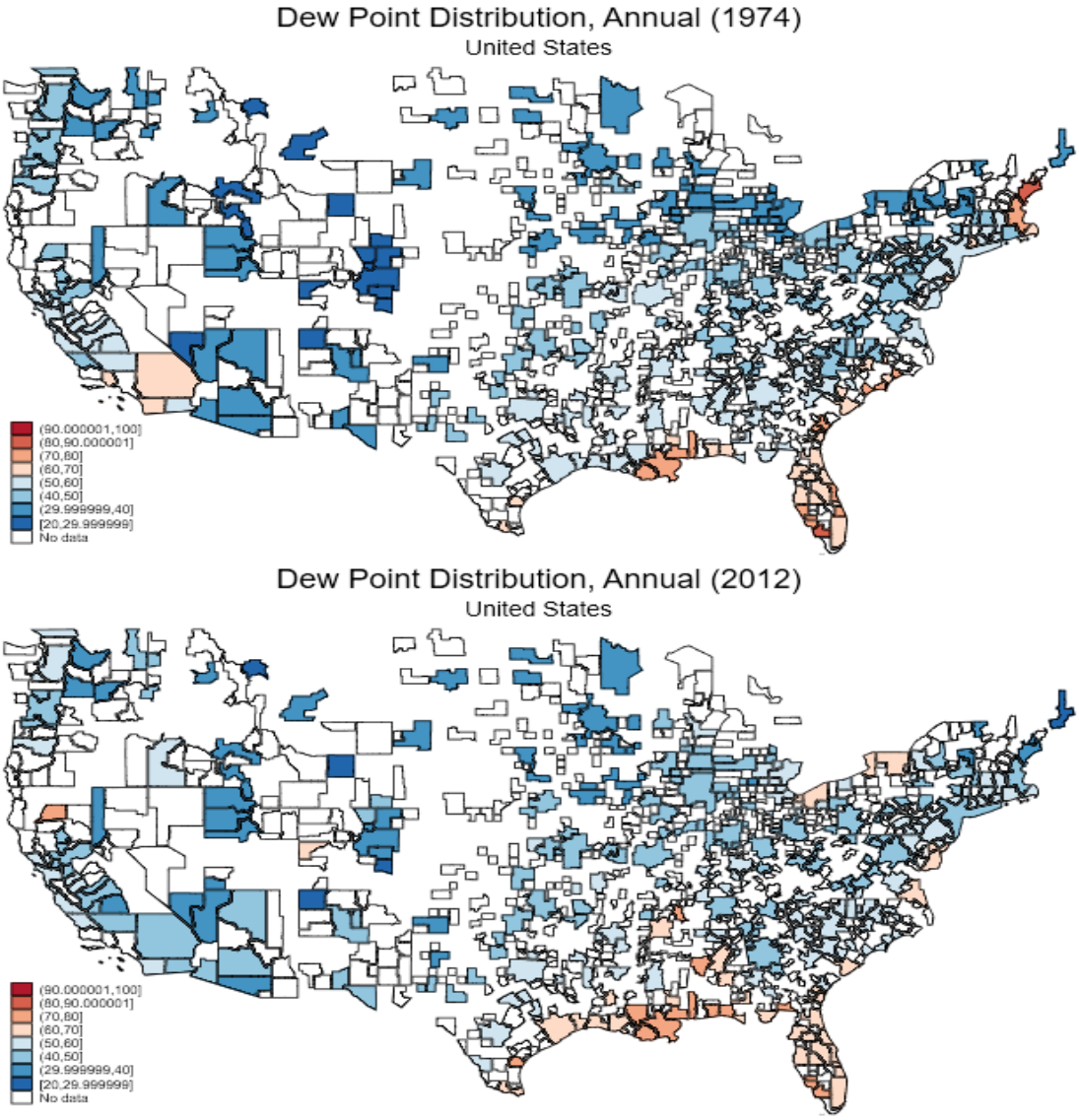


Figure A.3: GSOD data of Dew Point

Source: Author's Calculation using ArcGIS and GSOD

Appendix B

Appendix for Chapter II

B.1 Additional tables

Below are the supplementary tables for chapter II. Tables B.1 to B.3 report estimated MM coefficients. And B.4 reports applied restriction by state. All states have taken COVID-19 actions, but restrictions vary, and the length of time the measures are in place vary.

Table B.1: Estimated MM coefficients of a logistic model.

		1	2	3	4	5
$\vec{\Gamma}$	Residential Compactness	-0.43	0.02	0.02	0.02	0.02
		(0.031)	(0.003)	(0.003)	(0.003)	(0.003)
	Local self dependence	-0.01	0.00	0.00	0.00	0.00
		(0.002)	(0.001)	(0.001)	(0.001)	(0.001)
	% Working from Home	0.02	0.00	0.00	0.00	0.00
		(0.001)	(0.000)	(0.000)	(0.000)	(0.000)
	Political preference (% Democrat)	0.02	0.11	0.12	0.12	0.12
		(0.002)	(0.007)	(0.007)	(0.007)	(0.007)
	Health (More is poorer health)	0.01	0.00	0.00	0.00	0.00
		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
	Union Rights	0.00	0.04	0.04	0.03	0.03
		(0.005)	(0.005)	(0.005)	(0.006)	(0.005)
$\vec{G}$	Total cases in previous week	13.72	10.08	9.50	9.24	8.93
		(0.739)	(0.898)	(0.913)	(0.922)	(0.912)
	Neighboring effect	0.15	0.18	0.18	0.19	0.18
		(0.019)	(0.021)	(0.022)	(0.022)	(0.022)
R	Restriction's Index	0.16	0.15	0.15	0.15	0.15
		(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
$\vec{I}$	(Residential Compactness)^2	-19.17	-47.86	-53.67	-52.08	-48.79
		(1.940)	(2.665)	(2.658)	(2.616)	(2.692)
	(Local self dependence)^2	-0.01	0.04	0.04	0.04	0.04
		(0.007)	(0.006)	(0.006)	(0.006)	(0.006)
	(% Working from Home)^2	-0.04	0.01	0.02	0.02	0.02
		(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
	(% Democrat)^2	-2.86	0.92	1.86	1.58	1.05
		(0.454)	(0.489)	(0.489)	(0.491)	(0.503)
	(Health)^2	3.18	3.52	4.46	4.36	4.62
		(0.163)	(0.160)	(0.169)	(0.171)	(0.171)

Standard errors are in parentheses.

Table B.2: Estimated MM coefficients of a logistic model.

			1	2	3	4	5
$\vec{X}$	Age	Employed	-0.013	-0.032	-0.018	-0.020	-0.025
			(0.0057)	(0.0100)	(0.0098)	(0.0101)	(0.0103)
		Member		0.001	0.000	0.000	0.001
				(0.0023)	(0.0023)	(0.0023)	(0.0023)
		Covered		0.001	0.000	0.000	-0.001
				(0.0008)	(0.0008)	(0.0008)	(0.0009)
	Education	Employed	-0.004	1.249	0.788	0.817	0.653
			(0.1421)	(0.1710)	(0.1764)	(0.1783)	(0.1775)
		Member		-0.332	-0.267	-0.264	-0.260
				(0.0297)	(0.0293)	(0.0300)	(0.0300)
		Covered		0.140	0.114	0.121	0.125
				(0.0157)	(0.0157)	(0.0156)	(0.0158)
	Family Income	Employed	-0.011	-0.019	-0.022	-0.022	-0.023
			(0.0017)	(0.0020)	(0.0020)	(0.0020)	(0.0021)
		Member		0.002	0.002	0.002	0.002
				(0.0006)	(0.0006)	(0.0006)	(0.0006)
		Covered		-0.001	-0.001	-0.001	-0.001
				(0.0003)	(0.0003)	(0.0003)	(0.0003)
	% living in Metro area	Employed	-0.004	-0.006	-0.006	-0.006	-0.006
			(0.0011)	(0.0012)	(0.0012)	(0.0013)	(0.0013)
		Member		-0.003	-0.003	-0.003	-0.003
				(0.0007)	(0.0007)	(0.0007)	(0.0007)
		Covered		-0.001	-0.001	-0.001	-0.001
				(0.0003)	(0.0003)	(0.0003)	(0.0003)
	% Married	Employed	-0.020	0.015	0.016	0.015	0.018
			(0.0038)	(0.0035)	(0.0035)	(0.0035)	(0.0034)
		Member		0.000	0.000	0.000	0.000
				(0.0006)	(0.0006)	(0.0006)	(0.0006)
		Covered		0.001	0.001	0.001	0.001
				(0.0002)	(0.0002)	(0.0002)	(0.0002)
	% Male	Employed	0.121	0.112	0.121	0.120	0.122
			(0.0154)	(0.0071)	(0.0072)	(0.0072)	(0.0072)
		Member		0.001	0.001	0.002	0.002
				(0.0006)	(0.0006)	(0.0006)	(0.0006)
		Covered		0.000	0.000	0.001	0.000
				(0.0003)	(0.0003)	(0.0003)	(0.0003)
	% Public Sector		0.038	0.048	0.032	0.030	0.030
			(0.0052)	(0.0059)	(0.0061)	(0.0061)	(0.0061)

Specification 1 does not include socioeconomic characters for union members and covered nonmember employees. This is because it is supposed to be a predictive model, where there is no need to differentiate the treatment groups. Standard errors are in parentheses.

Table B.3: Estimated MM coefficients of a logistic model.

			1	2	3	4	5
$\vec{\Sigma}$	Management	Employed		0.117	0.250	0.182	0.089
				(0.1154)	(0.1171)	(0.1198)	(0.1236)
		Member		-0.046	-0.150	-0.154	-0.156
				(0.0829)	(0.0826)	(0.0835)	(0.0847)
		Covered		0.934	0.944	0.933	0.964
				(0.1460)	(0.1405)	(0.1370)	(0.1382)
	Professional	Employed		0.219	0.149	0.163	0.165
				(0.1061)	(0.1029)	(0.1059)	(0.1068)
		Member		0.062	-0.059	-0.069	-0.074
				(0.0870)	(0.0900)	(0.0887)	(0.0898)
		Covered		0.231	0.176	0.163	0.216
				(0.0969)	(0.0947)	(0.0946)	(0.0975)
	Nursing	Employed		0.166	0.159	0.149	0.172
				(0.0529)	(0.0530)	(0.0530)	(0.0534)
		Member		-0.049	-0.014	-0.022	-0.012
				(0.0506)	(0.0510)	(0.0498)	(0.0501)
		Covered		0.273	0.265	0.257	0.255
				(0.0681)	(0.0679)	(0.0686)	(0.0685)
	Service	Employed		0.292	0.297	0.295	0.321
				(0.0494)	(0.0486)	(0.0480)	(0.0489)
		Member		-0.015	-0.073	-0.093	-0.078
				(0.0491)	(0.0483)	(0.0482)	(0.0485)
		Covered		0.194	0.169	0.165	0.209
				(0.0583)	(0.0561)	(0.0562)	(0.0579)
	Sales & Office	Employed		-2.403	-1.150	-0.885	-0.942
				(0.6133)	(0.5889)	(0.6078)	(0.6121)
		Member		5.118	7.657	7.716	8.659
				(0.6288)	(0.6815)	(0.6863)	(0.7213)
		Covered		2.685	5.428	5.740	6.423
				(0.6932)	(0.7260)	(0.7475)	(0.7607)
$\vec{T}$	% Total	Employed			0.035	0.034	0.033
					(0.0037)	(0.0038)	(0.0038)
		Member		-0.083	-0.081	-0.092	-0.070
				(0.0168)	(0.0166)	(0.0165)	(0.0173)
		Covered		-0.112	-0.187	-0.178	-0.168
				(0.0623)	(0.0626)	(0.0616)	(0.0609)
	Hours	Employed					0.065
							(0.0130)
		Member				-0.007	-0.009
						(0.0022)	(0.0022)
		Covered				-0.002	-0.002
						(0.0012)	(0.0012)

Specification 1 does not include occupation shares. In specifications 2 to 5, we include occupation shares for all employees to prevent contaminating the results due to the different exposure to the COVID-19 for union members and/or covered nonmember employees. Standard errors are in parentheses.

Table B.4: Applied restrictions in states and DC.

State	Mass gathering restrictions	Initial business closure	Educational facilities closed	Non-essential services closed	Stay at home order	3/5 Restrictions appliead
Delaware	16-Mar-20	16-Mar-20	16-Mar-20	24-Mar-20	24-Mar-20	16-Mar-20
Maryland	16-Mar-20	16-Mar-20	16-Mar-20	23-Mar-20	30-Mar-20	16-Mar-20
Michigan	13-Mar-20	16-Mar-20	16-Mar-20	23-Mar-20	24-Mar-20	16-Mar-20
New Mexico	12-Mar-20	16-Mar-20	13-Mar-20	24-Mar-20	N/I	16-Mar-20
Washington	11-Mar-20	16-Mar-20	13-Mar-20	25-Mar-20	23-Mar-20	16-Mar-20
Connecticut	12-Mar-20	16-Mar-20	17-Mar-20	23-Mar-20	N/I	17-Mar-20
DC	13-Mar-20	17-Mar-20	16-Mar-20	25-Mar-20	30-Mar-20	17-Mar-20
Louisiana	13-Mar-20	17-Mar-20	16-Mar-20	22-Mar-20	23-Mar-20	17-Mar-20
Ohio	12-Mar-20	17-Mar-20	16-Mar-20	23-Mar-20	23-Mar-20	17-Mar-20
New Jersey	16-Mar-20	16-Mar-20	18-Mar-20	21-Mar-20	21-Mar-20	18-Mar-20
New York	12-Mar-20	16-Mar-20	18-Mar-20	22-Mar-20	22-Mar-20	18-Mar-20
Wisconsin	17-Mar-20	17-Mar-20	18-Mar-20	25-Mar-20	25-Mar-20	18-Mar-20
California	11-Mar-20	19-Mar-20	19-Mar-20	19-Mar-20	19-Mar-20	19-Mar-20
Hawaii	16-Mar-20	17-Mar-20	19-Mar-20	25-Mar-20	25-Mar-20	19-Mar-20
Indiana	12-Mar-20	16-Mar-20	19-Mar-20	24-Mar-20	25-Mar-20	19-Mar-20
Utah	17-Mar-20	19-Mar-20	16-Mar-20	N/I	N/I	19-Mar-20
Alabama	20-Mar-20	20-Mar-20	19-Mar-20	28-Mar-20	4-Apr-20	20-Mar-20
Kentucky	19-Mar-20	18-Mar-20	20-Mar-20	26-Mar-20	N/I	20-Mar-20
Wyoming	20-Mar-20	19-Mar-20	19-Mar-20	N/I	N/I	20-Mar-20
Illinois	13-Mar-20	21-Mar-20	17-Mar-20	21-Mar-20	21-Mar-20	21-Mar-20
Nevada	19-Mar-20	21-Mar-20	16-Mar-20	21-Mar-20	31-Mar-20	21-Mar-20
Texas	21-Mar-20	21-Mar-20	19-Mar-20	N/I	2-Apr-20	21-Mar-20
Colorado	19-Mar-20	17-Mar-20	23-Mar-20	26-Mar-20	26-Mar-20	23-Mar-20
Massachusetts	13-Mar-20	23-Mar-20	17-Mar-20	24-Mar-20	N/I	23-Mar-20
Missouri	23-Mar-20	17-Mar-20	23-Mar-20	N/I	6-Apr-20	23-Mar-20
Oregon	12-Mar-20	24-Mar-20	16-Mar-20	N/I	23-Mar-20	23-Mar-20
Pennsylvania	1-Apr-20	23-Mar-20	17-Mar-20	23-Mar-20	1-Apr-20	23-Mar-20
Rhode Island	17-Mar-20	23-Mar-20	16-Mar-20	N/I	28-Mar-20	23-Mar-20
Tennessee	23-Mar-20	23-Mar-20	20-Mar-20	1-Apr-20	2-Apr-20	23-Mar-20
Vermont	13-Mar-20	23-Mar-20	18-Mar-20	25-Mar-20	24-Mar-20	23-Mar-20
Alaska	24-Mar-20	17-Mar-20	16-Mar-20	28-Mar-20	28-Mar-20	24-Mar-20
Georgia	24-Mar-20	24-Mar-20	18-Mar-20	N/I	3-Apr-20	24-Mar-20
Montana	24-Mar-20	20-Mar-20	15-Mar-20	26-Mar-20	26-Mar-20	24-Mar-20
Virginia	15-Mar-20	24-Mar-20	16-Mar-20	N/I	30-Mar-20	24-Mar-20
West Virginia	24-Mar-20	18-Mar-20	14-Mar-20	24-Mar-20	25-Mar-20	24-Mar-20
Idaho	25-Mar-20	25-Mar-20	23-Mar-20	25-Mar-20	25-Mar-20	25-Mar-20
Maine	18-Mar-20	25-Mar-20	16-Mar-20	25-Mar-20	2-Apr-20	25-Mar-20
North Carolina	14-Mar-20	25-Mar-20	14-Mar-20	30-Mar-20	30-Mar-20	25-Mar-20
Oklahoma	24-Mar-20	25-Mar-20	17-Mar-20	1-Apr-20	N/I	25-Mar-20
Arkansas	27-Mar-20	19-Mar-20	17-Mar-20	N/I	N/I	27-Mar-20
Minnesota	27-Mar-20	17-Mar-20	18-Mar-20	N/I	27-Mar-20	27-Mar-20
New Hampshire	16-Mar-20	28-Mar-20	16-Mar-20	28-Mar-20	27-Mar-20	27-Mar-20
Arizona	30-Mar-20	20-Mar-20	16-Mar-20	N/I	30-Mar-20	30-Mar-20
Kansas	17-Mar-20	N/I	17-Mar-20	N/I	30-Mar-20	30-Mar-20
South Carolina	18-Mar-20	1-Apr-20	16-Mar-20	N/I	7-Apr-20	1-Apr-20
Nebraska	16-Mar-20	19-Mar-20	2-Apr-20	N/I	N/I	2-Apr-20
Florida	3-Apr-20	17-Mar-20	17-Mar-20	N/I	3-Apr-20	3-Apr-20
Mississippi	24-Mar-20	3-Apr-20	19-Mar-20	3-Apr-20	3-Apr-20	3-Apr-20
Iowa	17-Mar-20	17-Mar-20	4-Apr-20	N/I	N/I	4-Apr-20
North Dakota	N/I	20-Mar-20	16-Mar-20	N/I	N/I	
South Dakota	6-Apr-20	N/I	16-Mar-20	N/I	N/I	

N/I is *Not in Effect*

Bibliography

Acemoglu, D., Autor, D. H., and Lyle, D. (2004). Women, war, and wages: The effect of female labor supply on the wage structure at midcentury. *Journal of Political Economy*, 112(3):497–551.

Alderson, A. S., Beckfield, J., and Nielsen, F. (2005). Exactly how has income inequality changed? patterns of distributional change in core societies. *International Journal of Comparative Sociology*, 46(5-6):405–423.

Alderson, A. S. and Nielsen, F. (2002). Globalization and the great u-turn: Income inequality trends in 16 oecd countries. *American Journal of Sociology*, 107(5):1244–1299.

Allen, E. and Benfield, F. K. (2003). Environmental characteristics of smart growth neighborhoods, phase ii: Two nashville neighborhoods. *Natural Resources Defense Council, New York,¡ http://www. nrdc. org/cities/smartgrowth/char/charnash. pdf¿(Aug. 2010).*

Alonso, W. (1964). *Location and land use. Toward a general theory of land rent.* Harvard University Press Cambridge, MA.

American Community Survey (2018). Means of Transportation to Work. Table B08006.

Angel, S., Sheppard, S., Civco, D. L., Buckley, R., Chabaeva, A., Gitlin, L., Kraley, A., Parent, J., and Perlin, M. (2005). *The dynamics of global urban expansion.* Citeseer.

Asher, H. B. (2001). *American labor unions in the electoral arena.* Rowman & Littlefield Publishers.

Atkinson, A. B. (2003). Income inequality in oecd countries: Data and explanations. *CESifo Economic Studies*, 49(4):479–513.

Baum-Snow, N. (2007). Did highways cause suburbanization? *The Quarterly Journal of Economics*, 122(2):775–805.

Bhatta, B. (2010). *Analysis of urban growth and sprawl from remote sensing data.* Springer Science & Business Media.

Black, J. T. (1996). The economics of sprawl. *Urban Land*, 55(3):6–52.

Bloom, D. E. and Freeman, R. B. (1992). The fall in private pension coverage in the us. Technical report, National Bureau of Economic Research.

Boxall, P. and Purcell, J. (2011). *Strategy and human resource management.* Macmillan International Higher Education.

Brown, A., Collins, C., Frank, T., Haddow, K., Hitchings, B., Parry, S., Vanderpool, G., and Wormser, L. (1998). Sprawl: The dark side of the american dream. *Sierra Club, San Francisco, CA.*

Buchmueller, T. C., DiNardo, J., and Valletta, R. G. (2002). Union effects on health insurance provision and coverage in the united states. *ILR Review*, 55(4):610–627.

Budd, J. W. and Na, I.-G. (2000). The union membership wage premium for employees covered by collective bargaining agreements. *Journal of labor Economics*, 18(4):783–807.

Burchfield, M., Overman, H. G., Puga, D., and Turner, M. A. (2006). Causes of sprawl: A portrait from space. *The Quarterly Journal of Economics*, 121(2):587–633.

Clawson, M. (1962). Urban sprawl and speculation in suburban land. *Land economics*, 38(2):99–111.

Coutts, A. M., Beringer, J., and Tapper, N. J. (2007). Impact of increasing urban density on local climate: Spatial and temporal variations in the surface energy balance in melbourne, australia. *Journal of Applied Meteorology and Climatology*, 46(4):477–493.

Duany, A., Plater-Zyberk, E., and Speck, J. (2001). *Suburban nation: The rise of sprawl and the decline of the American dream*. Macmillan.

Duncan, G. M. and Leigh, D. E. (1985). The endogeneity of union status: An empirical test. *Journal of Labor Economics*, 3(3):385–402.

Duranton, G., Morrow, P. M., and Turner, M. A. (2014). Roads and trade: Evidence from the us. *Review of Economic Studies*, 81(2):681–724.

Ewing, R. (1997). Is los angeles-style sprawl desirable? *Journal of the American Planning Association*, 63(1):107–126.

Ewing, R., Haliyur, P., and Page, G. W. (1994). Getting around a traditional city, a suburban planned unit development, and everything in between. *Transportation Research Record*, 1466:53.

Ewing, R. and Hamidi, S. (2014). Measuring sprawl 2014. *Washington, DC: Smart Growth America. Accessed at from http://www. smartgrowthamerica. org/documents/measuring-sprawl-2014-pdf*, 8:2014.

Ewing, R., Pendall, R., and Chen, D. (2003). Measuring sprawl and its transportation impacts. *Transportation Research Record: Journal of the Transportation Research Board*, (1831):175–183.

Falcone, J. A. (2015). Us conterminous wall-to-wall anthropogenic land use trends (nwalt), 1974–2012. Technical report, US Geological Survey.

Flanagan, R. J. (1983). Workplace public goods and union organizations. *Industrial Relations: A Journal of Economy and Society*, 22(2):224–237.

Freeman, L. (2001). The effects of sprawl on neighborhood social ties: An explanatory analysis. *Journal of the American Planning Association*, 67(1):69–77.

Freeman, R. B. (1981). The effect of unionism on fringe benefits. *ILR Review*, 34(4):489–509.

Freeman, R. B. and Medoff, J. L. (1984). What do unions do. *Indus. & Lab. Rel. Rev.*, 38:244.

Galenson, W. (1986). The historical role of american trade unionism. *Unions in transition: Entering the second century*, pages 39–73.

Galster, G., Hanson, R., Ratcliffe, M. R., Wolman, H., Coleman, S., and Freihage, J. (2001). Wrestling sprawl to the ground: defining and measuring an elusive concept. *Housing policy debate*, 12(4):681–717.

Glaeser, E. L. and Kahn, M. E. (2004). Sprawl and urban growth. In *Handbook of Regional and Urban Economics*, volume 4, pages 2481–2527. Elsevier.

Gordon, P. and Richardson, H. W. (1997). Are compact cities a desirable planning goal? *Journal of the American planning association*, 63(1):95–106.

Green, R. K., Malpezzi, S., and Mayo, S. K. (2005). Metropolitan-specific estimates of the price elasticity of supply of housing, and their sources. *American economic review*, 95(2):334–339.

Harvey, R. O. and Clark, W. A. (1965). The nature and economics of urban sprawl. *Land Economics*, 41(1):1–9.

Hirsch, B. T. and Macpherson, D. A. (2003). Union membership and coverage database from the current population survey: Note (updated annually at unionstats.com). *ILR Review*, 56(2):349–354.

Institute for Health Metrics and Evaluation (2020). Covid-19 mortality, infection, testing, hospital resource use, and social distancing projections. Seattle, United States of America: Institute for Health Metrics and Evaluation (IHME), University of Washington.

Kaufman, B. E. (2004). What unions do: Insights from economic theory. *Journal of Labor Research*, 25(3):351–382.

Kermack, W. O. and McKendrick, A. G. (1927). A contribution to the mathematical theory of epidemics. *Proceedings of the royal society of london. Series A, Containing papers of a mathematical and physical character*, 115(772):700–721.

Malpezzi, S. et al. (1999). Estimates of the measurement and determinants of urban sprawl in us metropolitan areas. Technical report, University of Wisconsin Center for Urban Land Economic Research.

Martilli, A. (2014). An idealized study of city structure, urban climate, energy consumption, and air quality. *Urban Climate*, 10:430–446.

Mills, D. E. (1981). Growth, speculation and sprawl in a monocentric city. *Journal of Urban Economics*, 10(2):201–226.

Mishel, L., Bivens, J., Gould, E., and Shierholz, H. (2012). *The state of working America*. Cornell University Press.

MIT Election Data and Science Lab (2017). U.S. President 1976–2016.

Neal, T. (2013). Using panel co-integration methods to understand rising top income shares. *Economic Record*, 89(284):83–98.

Oke, T. R. (1973). City size and the urban heat island. *Atmospheric Environment (1967)*, 7(8):769–779.

Pierce, B. (1999). Using the national compensation survey to predict wage rates. *Compensation and Working Conditions*, 4(4):8–16.

Pizzella, P. and Beach, W. (2019). *National Compensation Survey: Employee Benefits in the United States.* Bulletin, 2791. U.S. Bureau of Labor Statistics, Washington, DC.

Rahimzadeh, G. (2020). The impact of urban sprawl on temperature in the united states over the past four decades. *Available at SSRN 3632565.*

Rones, P. L. (1981). Response to recession: reduce hours or jobs. *Monthly Lab. Rev.*, 104:3.

Schueler, T. (1994). The importance of imperviousness. *Watershed Protection Techniques*, 1(3):100–101.

Spilimbergo, A. (2009). Democracy and foreign education. *American economic review*, 99(1):528–43.

Stone, B., Hess, J. J., and Frumkin, H. (2010). Urban form and extreme heat events: are sprawling cities more vulnerable to climate change than compact cities? *Environmental Health Perspectives*, 118(10):1425.

Stone, B. and Rodgers, M. O. (2001). Urban form and thermal efficiency: how the design of cities influences the urban heat island effect. *Journal of the American Planning Association*, 67(2):186.

Strombom, B. A., Buchmueller, T. C., and Feldstein, P. J. (2002). Switching costs, price sensitivity and health plan choice. *Journal of Health economics*, 21(1):89–116.

The New York Times (2020a). Coronavirus (Covid-19) Data in the United States.

The New York Times (June 25 2020b). Only 1 in 10 U.S. coronavirus cases are likely to have been identified, C.D.C. chief says. Retrieved from https://www.nytimes.com/.

Weil, D. (1991). Enforcing osha: The role of labor unions. *Industrial Relations: A Journal of Economy and Society*, 30(1):20–36.

Weil, D. (1999). Are mandated health and safety committees substitutes for or supplements to labor unions? *ILR Review*, 52(3):339–360.

Weng, Q., Liu, H., and Lu, D. (2007). Assessing the effects of land use and land cover patterns on thermal conditions using landscape metrics in city of indianapolis, united states. *Urban Ecosystems*, 10(2):203–219.

Western, B. and Rosenfeld, J. (2011). Unions, norms, and the rise in us wage inequality. *American Sociological Review*, 76(4):513–537.

Wheaton, W. C. (2006). Metropolitan fragmentation, law enforcement effort and urban crime. *Journal of Urban Economics*, 60(1):1–14.

Yeh, A. (2001). Measurement and monitoring of urban sprawl in a rapidly growing region using entropy. *Photogrammetric Engineering & Remote Sensing*, 67(1):83–90.

Zhongming, Z., Wangqiang, Z., Wei, L., et al. (2017). Climate science special report: Fourth national climate assessment (nca4), volume i.

Vita

Golnoush Rahimzadeh was born in 1987 in Tehran, Iran. In 2005, she attended Allameh Tabatai University, Iran's highest ranking university in social science. She completed a BA with a major in Economics in 2009 and M.Sc. in Theoretical economics from the University of Tehran in 2012. During her years at UT, She developed her quantitative skills and computer literacy competency. She continued her education in the United States after being admitted into the East Carolina University (ECU) to study M.Sc. in Applied and Resource Economics. In the fall of 2014, Golnoush began her doctoral studies in Economics at Georgia State University. Golnoush has taught courses in Principles of Microeconomics and Econometrics. He has served as a research assistant working on a wide area of public policy. Her research interests are mainly directed toward Environmental Economics, Health, and Public Economics.